



# Guía de evaluación de la acción humanitaria

**ALNAP** es una red sectorial de membresía que promueve el aprendizaje activo y está empeñada en mejorar la calidad y la rendición de cuentas de las intervenciones humanitarias.

**<http://www.alnap.org>**

**Autores:**

Margie Buchanan-Smith  
John Cosgrave  
Alexandra Warner

**Con la colaboración adicional de:**

Francesca Bonino  
Neil Dillon  
Paul Knox Clarke  
Ralf Otto

Las opiniones vertidas en este informe no necesariamente reflejan los puntos de vista de las afiliadas de ALNAP.



ALNAP (2016), Guía de evaluación de la acción humanitaria  
Guía de ALNAP. Londres: ALNAP/ODI

© ALNAP/ODI 2016. Esta publicación ha sido autorizada  
bajo licencia Creative Commons/Licencia No Comercial  
(CC BY-NC 4.0).

ISBN 978-1-910454-49-7

Coordinación de comunicaciones:  
María Gili y Alex Glynn

Diseño: Human After All  
[www.humanafterall.co.uk](http://www.humanafterall.co.uk)

## Prólogo

**John Mitchell**, Director de ALNAP



Hace 20 años, cuando se creó ALNAP, la evaluación de la acción humanitaria aún se encontraba en un estado incipiente. Desde entonces la EAH ha evolucionado, hasta formar parte operativa del sistema humanitario en la actualidad, con incontables evaluaciones de buena calidad. El resultado es una mejora notable del aprendizaje y la rendición de cuentas en el sector humanitario.

El conocimiento generado a partir de la práctica de la EAH hoy forma parte de una base mundial de evidencia en crecimiento permanente, que está siendo usada para crear líneas de base a partir de las cuales monitorear el desempeño humanitario a través del tiempo, tal como lo demuestra el informe de ALNAP, El estado del sistema humanitario. La necesidad de material de evaluación de alta calidad con el fin de mejorar la precisión y el nivel de confianza de los informes de evaluación, será siempre vital para la comunidad humanitaria. Yo creo que esta guía brindará un gran servicio al respecto.

Desarrollar una guía verdaderamente exhaustiva ha sido una tarea monumental. Producirla ha tomado cinco años, ya que incide en múltiples temas vinculados a la evaluación humanitaria, todos los cuales forman parte de un entorno operativo sumamente complejo. La elaboración de la guía se ha beneficiado asimismo de un nivel de participación inusitadamente alto, con más de 40 organizaciones validando y suministrando comentarios durante el proceso piloto. Sin estos aportes, no hubiésemos podido crear una guía tan clara, pertinente y con tanta autoridad, que ya a estas alturas ha demostrado su valor para los evaluadores. Quisiera expresar mi más sincero agradecimiento a todos los que participaron en este esfuerzo.

Esperamos que la guía sea ampliamente utilizada y enriquezca la calidad y el disfrute de llevar a cabo evaluaciones de intervenciones humanitarias. Además, la guía seguirá fortaleciendo la rendición de cuentas y promoviendo una mejor evidencia evaluativa, conducente al aprendizaje de lecciones y al análisis, no solo a nivel del sistema en general sino también de organizaciones individuales y sus programas. Quisiéramos invitar a los usuarios a seguir suministrando retroalimentación sobre la guía, ya que estamos comprometidos a garantizar la calidad de la orientación, y por extensión el impacto de la EAH.



# Índice

## Introducción

03	Prólogo
10	Relación de siglas
12	Reconocimientos
14	Por qué es importante la evaluación
14	¿Por qué necesitamos una Guía de EAH?
16	Desarrollando la Guía de EAH de ALNAP
18	Cómo utilizar la Guía de EAH de ALNAP
20	¿Cómo está organizada la guía?
22	Código de los íconos incluidos en el texto

## 1 / ¿Qué es evaluación de la acción humanitaria?

24	1.1	Crisis humanitarias y acción humanitaria
26	1.2	Evaluación de la acción humanitaria
30	1.3	El seguimiento y la evaluación en el ciclo de respuesta a una emergencia
32	1.4	Desafíos comunes de la EAH

## 2 / La decisión de realizar una evaluación

41	2.1	¿Cuándo es apropiada una evaluación para la rendición de cuentas y el aprendizaje?
47	2.2	¿Es la evaluación la herramienta adecuada para la tarea?
48	2.3	Estableciendo un balance entre aprendizaje y rendición de cuentas
49	2.4	Decidiendo qué evaluar y cuándo
55	2.5	Ética y EAH

## 3/ Pensar desde el principio y a menudo en el uso de la evaluación

62	3.1	Qué significa estar enfocados en el uso
63	3.2	Cómo se utilizan las evaluaciones de acción humanitaria
65	3.3	Identificar, entender e involucrar a los usuarios previstos de la evaluación
72	3.4	El rol de la población afectada en la fase de planificación
73	3.5	Factores que afectan el uso de una evaluación

## 4 / Tipos de evaluación

- 78 4.1 ¿Qué tipo de evaluación van a emprender?
- 79 4.2 Definición de diferentes tipos de evaluación
- 87 4.3 Evaluaciones en tiempo real (ETR)
- 89 4.4 Evaluaciones conjuntas

## 5 / Enmarcando su evaluación

- 92 5.1 La lógica del programa
- 97 5.2 Teoría del cambio
- 100 5.3 Desarrollo de un modelo lógico retroactivo
- 101 5.4 Marcos de evaluación y criterios de evaluación

## 6 / Selección de las preguntas de evaluación

- 104 6.1 Impacto de la selección de las preguntas en la calidad de la evaluación
- 107 6.2 Tipos de preguntas de evaluación
- 109 6.3 Número de preguntas
- 110 6.4 Desentrañando las preguntas centrales
- 111 6.5 Criterios de evaluación
- 116 6.6 Selección de las preguntas más idóneas

## 7 / Términos de referencia y presupuesto de una evaluación

- 118 7.1 Términos de referencia
- 119 7.1.1 Los términos de referencia y el informe inicial
- 121 7.1.2 ¿Qué incluyen los TdR?
- 127 7.1.3 Línea de tiempo de los TdR
- 127 7.1.4 Extensión de los TdR
- 128 7.2 Elaboración del presupuesto de una evaluación

## 8 / Fase inicial

- 136 8.1 Informe inicial y control de calidad
- 137 8.2 Actividades de la fase inicial
- 138 8.3 La matriz de evaluación
- 142 8.4 El informe inicial

## 9 / Planificación y gestión de su evaluación

- 147 9.1 La decisión de evaluar
- 148 9.2 Grupos consultivos
- 153 9.3 Línea de tiempo de la evaluación
- 154 9.4 Evaluación interna y externa
- 157 9.5 Contratación de evaluadores
- 168 9.6 Desafíos para el liderazgo y el trabajo en equipo
- 170 9.7 Gestión de controversias

## 10 / Métodos de revisión documental

- 172 10.1 ¿Qué es una revisión documental?
- 174 10.2 ¿Por qué hacer una revisión documental?
- 175 10.3 ¿Cuánto tiempo debe durar una revisión documental?
- 175 10.4 Llevando a cabo una revisión documental
- 178 10.5 Herramientas para revisiones documentales
- 182 10.6 Resumiendo la información de la revisión documental

## 11 / Diseños de evaluación para responder las preguntas de evaluación

- 193 11.1 Cualitativos y cuantitativos
- 194 11.2 Familias de diseño
- 200 11.3 Selección del diseño
- 209 11.4 Sesgo

## 12 / Muestreo

- 215 12.1 Problemas muestrales comunes en la EAH
- 216 12.2 Muestreo no aleatorio
- 224 12.3 Muestreo aleatorio
- 230 12.4 Muestreo para métodos mixtos

## 13 / Metodologías de campo

- 232 13.1 La entrevista como técnica
- 236 13.2 Intérpretes
- 238 13.3 Métodos de entrevista cualitativa
- 246 13.4 Métodos de encuesta
- 253 13.5 Observación
- 257 13.6 Mediciones discretas
- 259 13.7 Metodologías orientadas al aprendizaje

## 14 / Involucrando a la población afectada en su evaluación

- 268 14.1 Involucrando a la población afectada
- 271 14.2 Planificando involucrar a la población afectada
- 273 14.3 Diseñando la evaluación de modo tal que la población afectada se vea involucrada
- 274 14.4 Métodos para involucrar a la población afectada
- 279 14.5 Desafíos particulares de involucrar a la población afectada en la EAH
- 280 14.6 Consideraciones éticas de involucrar a las poblaciones afectadas

## 15 / Acceso restringido

- 283 15.1 ¿Qué es el acceso restringido?
- 284 15.2 Importancia del análisis de evaluabilidad
- 286 15.3 Opciones para enfrentar el acceso restringido
- 291 15.4 Credibilidad de la evaluación a distancia
- 292 15.5 Otras alternativas a la evaluación a distancia

## 16 / Análisis

294	16.1 <i>Big-n o small-n</i>
296	16.2 Análisis de evidencia para responder las preguntas normativas
297	16.3 Análisis de evidencia para preguntas evaluativas
302	16.4 Análisis de evidencia para preguntas causales
303	16.5 Análisis de evidencia para responder las preguntas descriptivas
311	16.6 Análisis estadístico de información primaria
316	16.7 Análisis numérico de información secundaria
318	16.8 De la evidencia a las recomendaciones

## 17 / Reportando y comunicando los hallazgos de la evaluación con enfoque de utilización

328	17.1 Productos centrales
330	17.2 Informe de evaluación
339	17.3 Difusión
342	17.4 Diferentes canales de comunicación
346	17.5 Cómo facilitar la aceptación de una evaluación
349	17.6 Síntesis de evaluaciones, revisiones temáticas y metaevaluaciones para potenciar la aceptación

## 18 / Evaluaciones del impacto humanitario

356	18.1 ¿Por qué es importante la evaluación de impacto?
356	18.2 ¿Qué es el impacto en la EAH?
359	18.3 Planificación de una evaluación de impacto
362	18.4 Enfoques y diseños
367	Glosario

Ver la bibliografía y el índice en: [www.alnap.org/EHA](http://www.alnap.org/EHA).

## Relación de siglas

<b>ACF</b>	Acción contra el Hambre
<b>ACNUR</b>	Alto Comisionado de las Naciones Unidas para los Refugiados
<b>ACT</b>	Acción Conjunta de las Iglesias (Action by Churches Together)
<b>ALNAP</b>	Red de Aprendizaje Activo sobre Rendición de Cuentas y Resultados de la Acción Humanitaria (Active Learning Network for Accountability and Performance in Humanitarian Action)
<b>APS</b>	Agua potable, saneamiento e higiene
<b>CAQDAS</b>	Análisis computarizado de datos cualitativos
<b>CDA</b>	Proyectos de aprendizaje colaborativo de CDA
<b>CDAC</b>	Red de Comunicación con Comunidades Afectadas por Desastres
<b>CERF</b>	Fondo Central de Respuesta a Emergencias
<b>CFW</b>	Iniciativa “dinero por trabajo” (cash for work)
<b>CH</b>	Coordinador humanitario
<b>CHS</b>	Norma humanitaria esencial
<b>CICR</b>	Comité Internacional de la Cruz Roja
<b>CMS</b>	Cambio más significativo
<b>CRS</b>	Catholic Relief Services
<b>CwC</b>	Iniciativa “Communicating with Communities” (Comunicándonos con las Comunidades)
<b>DEC</b>	Comité de Emergencia para Desastres
<b>DFID</b>	Departamento (Ministerio) para el Desarrollo Internacional del Reino Unido
<b>DGF</b>	Discusión de grupo focal
<b>DI</b>	Desplazados/as internos/as
<b>DRP</b>	Diagnóstico rápido/rural participativo
<b>DRC</b>	Consejo Danés para los Refugiados
<b>EAH</b>	Evaluación de la acción humanitaria
<b>ECB</b>	Proyecto de Fortalecimiento de las Capacidades de Emergencia
<b>ECHO</b>	Dirección General de Ayuda Humanitaria y Protección Civil de la Comisión Europea
<b>ECIH</b>	Evaluación conjunta del impacto humanitario
<b>EHII</b>	Evaluación Humanitaria Interinstitucional
<b>EHP</b>	Equipo humanitario país
<b>ER</b>	Examen retrospectivo
<b>ETR</b>	Evaluación en tiempo real
<b>FAO</b>	Organización de las Naciones Unidas para la Alimentación y la Agricultura
<b>FFS</b>	Enfoque de las escuelas de campo para agricultores (Farmer Field School)
<b>FICR</b>	Federación Internacional de Sociedades de la Cruz Roja y la Media Luna Roja
<b>HAP</b>	Asociación Internacional para la Rendición de Cuentas en el Ámbito Humanitario (Humanitarian Accountability Partnership)

<b>IASC</b>	Comité Permanente Interagencial
<b>IRC</b>	Comité Internacional de Rescate
<b>JEEAR</b>	Evaluación conjunta de la asistencia a la emergencia de Ruanda
<b>LRRD</b>	Proyecto para vincular la ayuda de emergencia, la rehabilitación y el desarrollo
<b>MSF</b>	Médicos Sin Fronteras
<b>NRC</b>	Consejo Noruego para Refugiados
<b>OBC</b>	Organización de base comunitaria
<b>OCAH</b>	Oficina de Coordinación de los Asuntos Humanitarios de la ONU
<b>OCDE-CAD</b>	Organización para la Cooperación y el Desarrollo Económico – Comité de Asistencia al Desarrollo
<b>ODI</b>	Instituto del Desarrollo en Ultramar (Overseas Development Institute)
<b>OFDA</b>	Oficina de los Estados Unidos de Asistencia para Desastres (USAID)
<b>OMS</b>	Organización Mundial de la Salud
<b>ONG</b>	Organización No Gubernamental
<b>PDA/ADP</b>	Asistente digital personal
<b>PMA</b>	Programa Mundial de Alimentos
<b>PNUD</b>	Programa de las Naciones Unidas para el Desarrollo
<b>RAPID</b>	Programa de investigación y políticas del desarrollo del ODI (Research and Policy in Development Programme)
<b>RCA</b>	República Centroafricana
<b>RCPA</b>	Rendición de cuentas a la población afectada
<b>RCT</b>	Prueba controlada aleatoria (randomised control trial)
<b>RDC</b>	República Democrática del Congo
<b>RDC</b>	Reconstrucción dirigida por la comunidad
<b>ROI</b>	Iniciativa de la Región de Origen en Afganistán
<b>ROMA</b>	Enfoque de Mapeo de Efectos del programa RAPID
<b>RRD</b>	Reducción del riesgo de desastres
<b>SAVE</b>	Acceso Seguro en Entornos Volátiles
<b>SCHR</b>	Comité Directivo para la Respuesta Humanitaria
<b>SEA</b>	Seguimiento, evaluación y aprendizaje
<b>Sida</b>	Agencia Sueca de Cooperación para el Desarrollo Internacional
<b>SyE</b>	Seguimiento y evaluación
<b>TdC</b>	Teoría del Cambio
<b>TdR</b>	Términos de referencia
<b>TEC</b>	Coalición de Evaluación del Tsunami
<b>UNEG</b>	Grupo de Evaluación de las Naciones Unidas
<b>UNEG HEIG</b>	Grupo de Interés en la Evaluación Humanitaria del Grupo de Evaluación de las Naciones Unidas
<b>UNICEF</b>	Fondo de las Naciones Unidas para la Infancia
<b>WVI</b>	World Vision International

## Reconocimientos

La elaboración de la presente Guía de evaluación de la acción humanitaria se llevó a cabo en el marco de un amplio proceso participativo, facilitado por la apertura, intensa actividad y dedicación de un gran número de personas y organizaciones. La Secretaría de ALNAP quisiera agradecer especialmente a algunas de ellas.

En primer lugar, la guía no hubiera sido posible sin la participación positiva sostenida de las afiliadas de ALNAP. La Secretaría de ALNAP está profundamente agradecida por la honestidad y el entusiasmo de la membresía: sin su apoyo, la creación de un recurso práctico como este no hubiera sido posible.

La Secretaría de ALNAP quisiera agradecer asimismo a John Cosgrave y Margie Buchanan-Smith, por su compromiso permanente y su apasionada dedicación a la redacción, mejora, promoción y utilización de la Guía de EAH, así como de otros materiales de capacitación previos sobre EAH producidos en los últimos años.

Nuestro agradecimiento muy especial también a Wendy Fenton, de Humanitarian Practice Network (HPN), quien circuló, al inicio del proceso, una encuesta entre las afiliadas de HPN. Muchas de ellas reconocieron la necesidad de una nueva guía de evaluación de la acción humanitaria.

Agradecemos la valiosa retroalimentación de los integrantes del grupo consultivo formado por representantes sectoriales de todas las partes interesadas de la evaluación: Jock Baker, Tony Beck, Hana Crowe, Stefan Dahlgren, Wendy Fenton, Enrique García, Josse Gillijns, Saúl Guerrero, Babar Kabir, Peter Klansoe, Caroline Loftus, Rob McCouch, Joakim Molander, Jonathan Patrick, Nicoletta Pergolizzi, Riccardo Polastro, Hannah Reichardt, Peta Sandison y Vivien Walden. Sus comentarios ayudaron a la Secretaría de ALNAP con el proceso de redacción.

Un enorme “gracias” a los cientos de personas que participaron en el piloto, utilizaron la guía de EAH y se tomaron el tiempo de analizarla y hacernos llegar sus apreciaciones, comentarios y sugerencias de mejoras, o simplemente unas palabras de reconocimiento y apreciación. Nos impactó la disposición de las afiliadas de ALNAP, y no solamente de ellas, para participar en el piloto, evaluar el curso de aprendizaje en línea sobre EAH desarrollado conjuntamente con UNICEF y compartir sus puntos de vista sobre cómo potenciar la utilidad de la guía así como su focalización en los operadores de respuestas humanitarias. Su retroalimentación fue invaluable durante el proceso de revisión y finalización de la Guía de EAH.

Agradecemos al selecto grupo de participantes en el piloto y operadores de EAH que nos ayudaron a procesar algunos de los comentarios más difíciles y de más alto nivel del taller de validación al término del piloto. Nuestro agradecimiento

especial a: Sarah Bailey, Lian Bradley, Annie Devonport, Kathy Duryee, Josse Gillijns, Langdon Greenhalgh, H elene Juillard, Anne-Claire Luzot, Velina Mikova, Mikkel Nedergaard, Joanna Olsen, Juliet Parker, Koorosh Raffii, Anke Reiffenstuel y Bonaventure Sokpoh. Nuestro reconocimiento asimismo a Ralf Otto, por su moderaci n del debate y su representaci n del grupo como revisor de pares de la versi n final.

Nuestro sincero agradecimiento a Sian Cook, por su invaluable asistencia de investigaci n durante el proceso de revisi n de la Gu a de EAH.

Muchas gracias a John Mitchell y Paul Knox Clarke por su liderazgo y orientaci n permanentes, as  como al equipo de Comunicaciones de la Secretar a de ALNAP por su apoyo y creatividad constantes, que hicieron de la Gu a de EAH un recurso asequible, no intimidante. Gracias en particular a Patricia Curmi, Franziska Orphal, Maria Gili y Alex Glynn.

Por  ltimo, y ciertamente no por ello menos importante, nuestro profundo agradecimiento a Francesca Bonino, quien coordin  esta iniciativa por espacio de cuatro a os. Tu energ a y entusiasmo contribuyeron a propulsar este proyecto, que no hubiera sido posible sin tu esfuerzo.

## Por qué es importante la evaluación

¿Cómo saber si nuestros esfuerzos humanitarios están dando resultado?

La evaluación es un importante medio para averiguarlo. En esencia, el objetivo de la evaluación es realizar un juicio de valor informado sobre las actividades y sus resultados. ¿Hemos marcado una diferencia? ¿Ayudamos realmente a salvar vidas y aliviar el sufrimiento? ¿Lo hicimos de la mejor manera posible? Las evaluaciones sólidas, robustas, nos permiten emitir un juicio ponderado, basado en evidencia, sobre el nivel de éxito de un programa, así como sobre la naturaleza de dicho éxito. Permiten a las partes interesadas responder la pregunta “¿entonces qué?” (Morra Imas y Rist, 2009) – trascendiendo las afirmaciones generales referidas al éxito de un proyecto, programa o política para enfocarse en si valió la pena o no, por qué y para quién (Buffardi et al., 2015).

Presiones que se remontan a la década de los 90 para reforzar la calidad y la rendición de cuentas del sector humanitario (Knox-Clarke y Darcy, 2013), así como presiones más recientes para demostrar la relación precio-calidad y optimizar el uso de recursos ajustados (ALNAP, 2015), han elevado sustancialmente la importancia de evaluaciones de calidad en la acción humanitaria. Una evaluación bien planificada, diseñada y ejecutada es un proceso que puede contribuir tanto al aprendizaje como a la rendición de cuentas a distintos niveles. En contextos que demandan un ritmo de trabajo muchas veces frenético, las evaluaciones ofrecen una valiosa oportunidad para hacer un alto en el camino y realizar un balance objetivo de las cosas que están funcionando o han funcionado bien y las cosas que tienen que cambiar, contribuyendo con una mirada estructurada al desempeño general de un proyecto, programa o respuesta. Contribuyen al cuerpo de evidencia sobre las cosas que funcionan y no funcionan bien en estos contextos, que a menudo comportan grandes dificultades. Las evaluaciones sirven para responder por lo tanto a una difícil pregunta: “¿Cómo nos estamos desempeñando *realmente*?” – y ayudar a las instancias de decisión a efectuar los correctivos de rumbo necesarios o tomar decisiones difíciles.

## ¿Por qué necesitamos una Guía de EAH?

Desde la década de los 90 hemos visto un sinfín de evaluaciones de acciones humanitarias, las mismas que han generado una masa crítica de conocimiento y experiencias colectivas que pueden ser capitalizadas. En 2009, sin embargo, aún no existía una guía sectorial integral sobre la evaluación de intervenciones humanitarias. El trabajo en la guía empezó a instancias de las afiliadas de la Humanitarian Practice Network. Invitadas a elegir un solo material de orientación que necesitarían de manera indispensable, la respuesta fue inequívoca: una guía de EAH.

## Seis razones por las que hace falta una Guía de EAH:

1

El interés y las inversiones en evaluación aumentan día a día, a medida que surgen cuestionamientos respecto de la rendición de cuentas y la eficacia de la cooperación internacional y la acción humanitaria.

2

Actualmente existe una masa crítica de conocimiento colectivo que puede ser capitalizada – tan solo la base de datos de evaluación de ALNAP contiene más de 2.000 evaluaciones que se remontan a los últimos 30 años.

3

Es preciso crear un lenguaje y una comprensión comunes en materia de EAH, a fin de facilitar las discusiones tanto dentro de los equipos y organizaciones como entre organizaciones.

4

Si bien la práctica de la evaluación es ahora más común, relativamente pocas evaluaciones manejan conceptos claros respecto de la metodología. Los organismos y evaluadores humanitarios que sí lo hacen a menudo se circunscriben a un número reducido de diseños y metodologías. Existe un margen de mejora en cuanto al uso de los diseños y enfoques de evaluación y a la selección de la metodología más idónea para responder a las preguntas de evaluación formuladas.

5

Del mismo modo, las evaluaciones siguen siendo una herramienta común para las instituciones donantes, que desean evaluar la rendición de cuentas de las intervenciones que financian, pero también pueden ser aprovechadas por las organizaciones para extraer aprendizajes y mejorar sus programas. En general, siguen existiendo muchas oportunidades para que las organizaciones humanitarias aprovechen al máximo las evaluaciones que responden a estas necesidades.

6

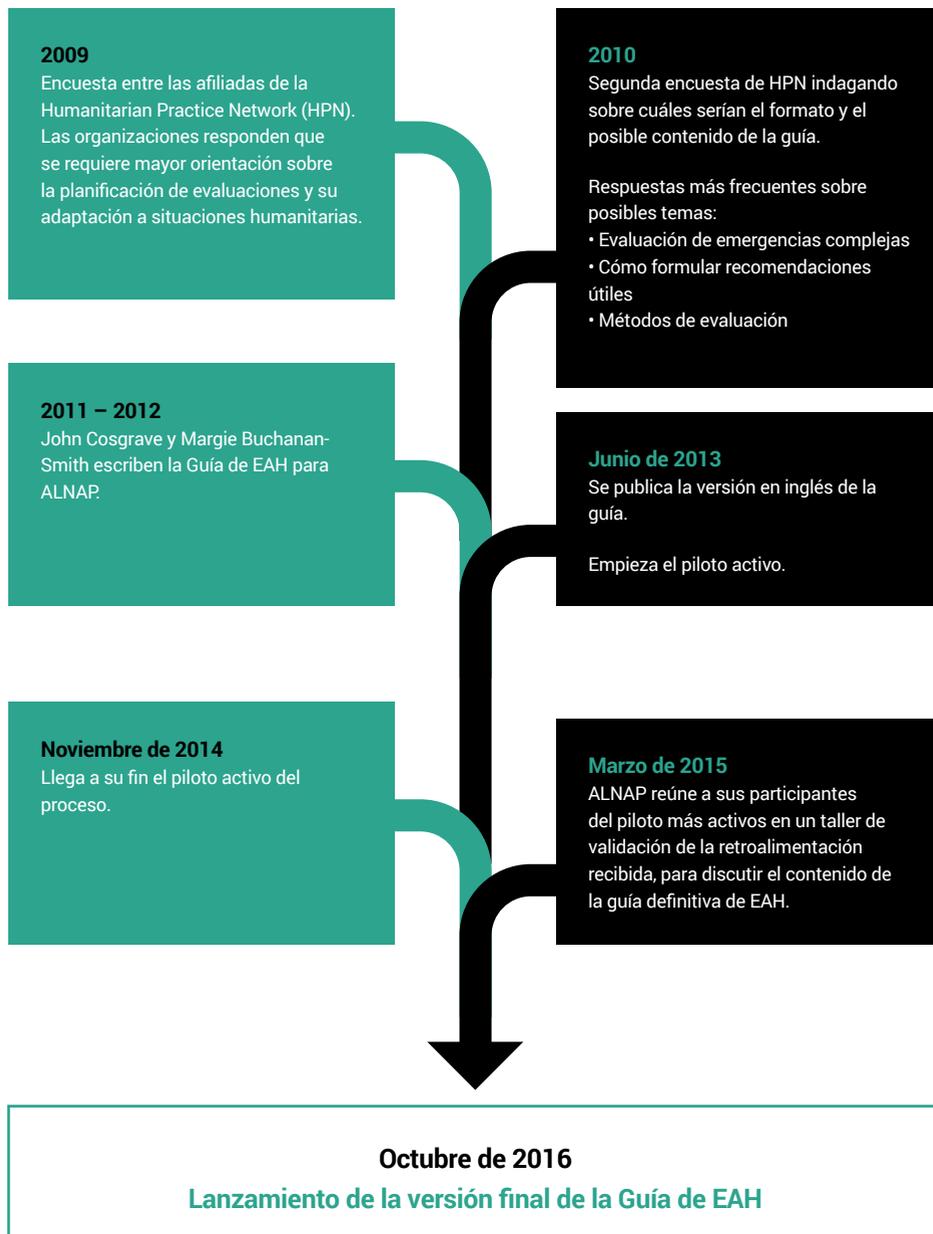
La contratación de evaluaciones se ha desplazado, de las oficinas centrales de las organizaciones al personal en terreno, a medida que las organizaciones inician un proceso de descentralización – y sin embargo los directores de las oficinas de terreno a menudo tienen poca experiencia en la planificación y gestión de evaluaciones, especialmente EAH.

## Desarrollando la Guía de EAH de ALNAP

El desarrollo de la guía de evaluación de la acción humanitaria de ALNAP se apoyó en un extenso proceso de investigación y consulta. La guía:

- Se basa en materiales de capacitación que ALNAP encargó hacer en 2001. Dichos materiales fueron validados, actualizados y adaptados en repetidas oportunidades a través de los años.
- Fue redactada en consulta con un grupo consultivo formado por representantes sectoriales de todas las partes interesadas de la evaluación.
- Fue activamente validada en el marco de un proceso piloto de 18 meses de duración. Más de 40 organizaciones humanitarias utilizaron la guía y ofrecieron comentarios sobre ella. El proceso culminó con un taller de validación de la retroalimentación recibida, que fue realizado al término del piloto, fue organizado por ALNAP y contó con la asistencia de 15 de los participantes en el piloto y operadores de EAH más activos. Este grupo ayudó a la Secretaría de ALNAP a procesar la retroalimentación de más alto nivel durante el proceso de revisión final.
- Fue revisada y actualizada durante la elaboración de la introducción de ALNAP y UNICEF al curso de EAH en línea, en colaboración con UNEG y EvalPartners. Las secciones revisadas de la guía se basan principalmente en el extenso proceso de revisión que se llevó a cabo para rediseñar la guía para la versión en línea.
- Fue revisada con base en la retroalimentación recibida del proceso piloto. Se revisaron informes de evaluación y otros recursos evaluativos con el fin de incorporar nuevos ejemplos.

## Línea de tiempo de la Guía de EAH



## Cómo utilizar la Guía de EAH de ALNAP

### ¿A quién está dirigida la guía?

La presente guía está diseñada para ayudar a todas las personas e instituciones que están pensando planificar, diseñar y ejecutar evaluaciones y/o pretenden utilizarlas, a capitalizar esa masa crítica de conocimiento, en especial la gran variedad de ejemplos de buenas prácticas. La guía ha sido creada para satisfacer una necesidad manifiesta y ayudar a las personas responsables de encargar, coordinar, diseñar, ejecutar y utilizar evaluaciones a sacar el mayor provecho posible del proceso de evaluación, y en última instancia a mejorar la calidad de la EAH, así como la forma en que esta es utilizada.

### Detalles a tener en cuenta al utilizar esta guía

El objetivo de la presente guía es impulsar evaluaciones de alta calidad, que contribuyan a mejorar el desempeño ofreciendo la mejor evidencia posible de las cosas que están y no están funcionando bien y por qué. Para ello, existen unas cuantas consideraciones que los lectores deben tomar en cuenta a la hora de utilizar la guía.

En primer lugar, y ante todo, las evaluaciones de las intervenciones humanitarias cuestan dinero, que de otra manera podría ser utilizado para prevenir pérdidas humanas o aliviar el sufrimiento. Este dinero habrá sido bien gastado si conduce a mejoras en la acción humanitaria, pero esto solo sucederá si los hallazgos de la evaluación son de alta calidad y son utilizados (Hallam y Bonino, 2013). Una evaluación no debe llevarse a cabo como una actividad aislada, sino que debe formar parte de los esfuerzos de una organización para rendir cuentas sobre su trabajo y aprender. Las evaluaciones tendrán mayor influencia en contribuir a mejorar el rendimiento de las organizaciones cuando estas, y el sector en general, se comprometan a averiguar cómo se están desempeñando y a extraer aprendizajes.

La EAH cumple por lo tanto una función que trasciende ampliamente la rendición de cuentas. Eso no quiere decir que las evaluaciones no sean una importante herramienta de rendición de cuentas – lo son. Pero rendición de cuentas implica más que simplemente escribir informes. Significa también interiorizar lo que dichos informes dicen y hacerles caso en aras de implementar mejoras. El aprendizaje es una parte importante de ello. Como señala Irene Guijt: “No se puede rendir cuentas si no se aprende” (2010: 277). Muchas veces, las evaluaciones son percibidas, y hasta temidas, como una forma de crítica o solo como un requisito de las entidades donantes. Pero pueden y deben jugar un papel mucho más constructivo y útil. Al adaptar las evaluaciones a las necesidades de sus usuarios primarios previstos, se convierten en oportunidades fundamentales para aprender y mejorar los programas. Esta guía ha sido escrita con un fuerte énfasis en el uso.

Muchos de los desafíos y complejidades a los que se enfrenta la acción humanitaria son un reflejo de los retos que afronta el trabajo de la evaluación, siendo quizás la falta de acceso a las poblaciones afectadas y a los lugares afectados una de las características más difíciles (Hallam y Bonino, 2013: 12). De modo que ¿cómo llevar a cabo evaluaciones lo suficientemente rigurosas y verosímiles en dichos contextos? Responder a esta pregunta representa el eje vertical de esta guía, y se aborda en todas las secciones.

La Guía de EAH considera las etapas estructuradas que debe tener una evaluación para ayudarles a obtener resultados de calidad. Cada una de estas etapas debe ser adaptada no solo al contexto de la evaluación (teniendo en cuenta factores tales como: crisis, respuesta, país, región, proyecto o programa, o equipo) sino también al contexto institucional (por ejemplo, liderazgo, cultura de aprendizaje). La guía presenta ejemplos extraídos de evaluaciones tanto pequeñas como grandes. Es importante establecer el nivel de aspiración de una EAH a la luz de los recursos disponibles, en términos tanto humanos como financieros. Para ayudar con ello, la Guía de EAH ofrece algunos criterios para evaluaciones más pequeñas.

La guía aprovecha las investigaciones que ya han sido realizadas y plasmadas en una serie de publicaciones y recursos de ALNAP pero evita duplicar dichas informaciones. Si bien aspira a ser exhaustiva, sin duda habrá algunos temas de actualidad que no han sido cubiertos, y probablemente emerjan temas adicionales a medida que evoluciona la acción humanitaria. Un ejemplo es la evaluación de actividades de protección humanitaria. La red de ALNAP está preparando una guía de evaluación de iniciativas de protección, que será publicada como acompañamiento de esta guía. Con el tiempo pueden surgir otros temas nuevos que serán abordados igualmente en guías conexas.

## ¿Cómo está organizada la guía?

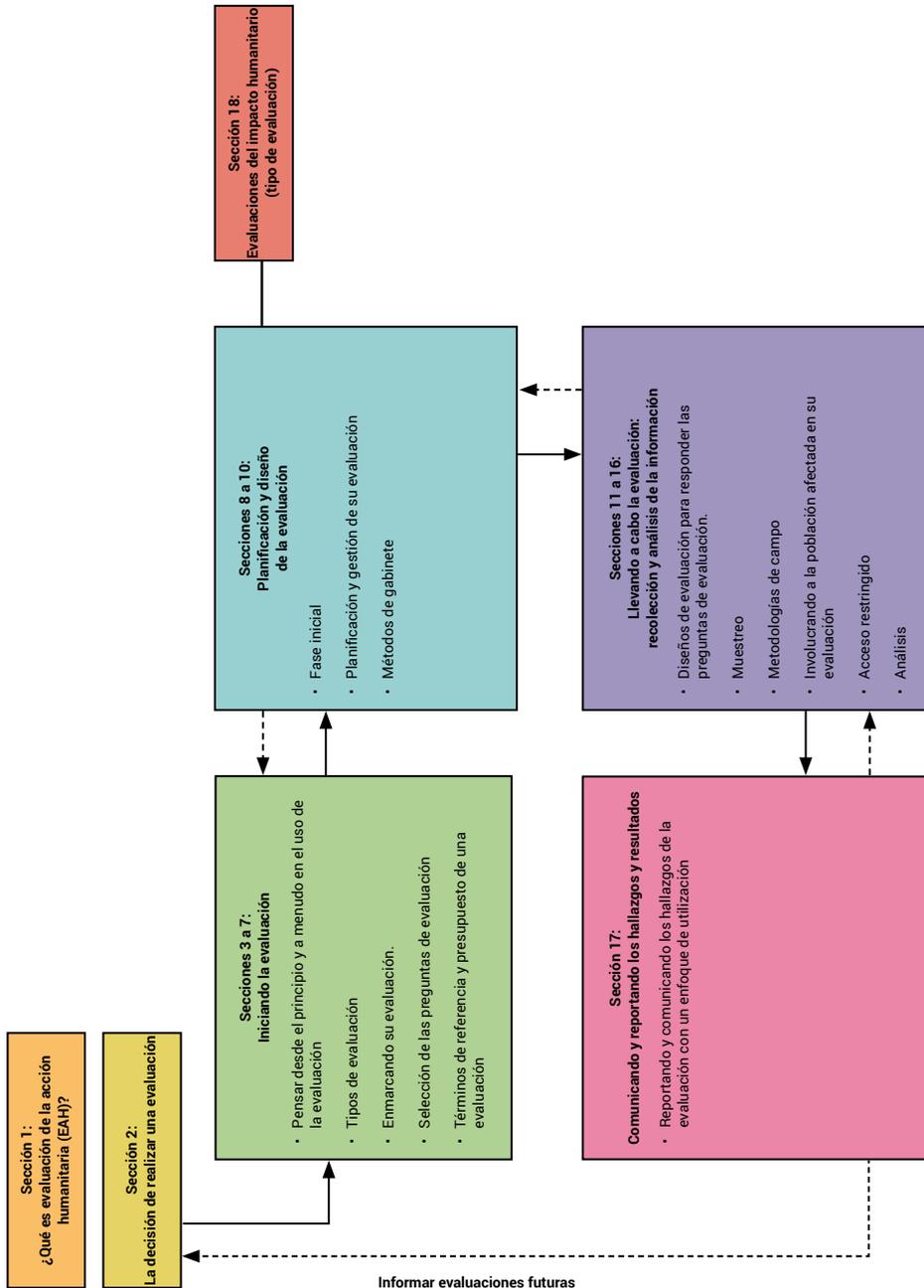
La Guía de EAH está organizada de tal manera que refleja un proceso de evaluación típico. Acompaña al usuario a través de las etapas que la organización que encarga la evaluación, el gestor de la evaluación y el evaluador acometerían desde que se toma la decisión de realizar una evaluación hasta la difusión de los resultados. La guía está dividida en cinco capítulos que reflejan las etapas importantes del proceso de evaluación, subdividiéndose nuevamente en secciones más pequeñas que inciden en las actividades específicas que se llevan a cabo en cada una de estas etapas. Las etapas y secciones se encuentran graficadas en el mapa que aparece más adelante. Dicho mapa puede servir como una herramienta para navegar a través de la guía pero ilustra asimismo el proceso de evaluación completo con bucles de retroalimentación.

Los procesos de evaluación de la calidad no son lineales. En primer lugar, la Guía de EAH hace hincapié en la importancia de enfocarse en el uso. Los resultados de una evaluación deben alimentar el trabajo humanitario futuro, al servir como insumos para la toma de decisiones, facilitar el aprendizaje, identificar las áreas susceptibles de mejora y conducir a otras preguntas de investigación o evaluación. Para reflejar lo anterior, las diferentes etapas de evaluación se representan en forma de bucles poco rígidos, antes que en una línea de tiempo.

En segundo lugar, los procesos de evaluación tienen una serie de bucles de retroalimentación, y puede haber algunas idas y vueltas entre las actividades. En la guía, estos vínculos entre secciones se resaltan por medio de referencias cruzadas.

Ver la bibliografía y el índice en: [www.alnap.org/EHA](http://www.alnap.org/EHA).

# Mapa de la Guía de EAH: Cómo navegar las diferentes etapas y secciones



## Código de los íconos incluidos en el texto

### Íconos



Preguntas clave



Definición



En profundidad



Ejemplo de buena práctica



Tener presente



Consejo útil

### Cuadros



Los cuadros del color de la sección deben leerse como parte del contenido de la guía.



Los cuadros de color gris son ejemplos tomados de la vida real que apoyan el texto.

### Diagramas



Directamente



Indirectamente

# ¿Qué es evaluación de la acción humanitaria?



# 1 / ¿Qué es evaluación de la acción humanitaria?

## 1.1 Crisis humanitarias y acción humanitaria

Las crisis humanitarias se desatan por numerosos motivos y pueden tener características muy diversas. He aquí algunas de las formas más comunes de clasificarlas y describirlas:

1. **Causa:** por ejemplo, conflicto armado; desastre natural de ocurrencia rápida como terremoto o inundación; desastre natural de ocurrencia lenta como sequía; crisis de salud como el ébola (FICR, sin fecha);
2. **Duración y frecuencia:** por ejemplo, prolongada, recurrente, de ocurrencia rápida o lenta (ALNAP, 2007);
3. **Zona afectada:** por ejemplo, urbana, rural, megacrisis (FICR, 2010);
4. **Capacidad de respuesta estatal y/o nacional:** por ejemplo, si la capacidad estatal es reducida, las organizaciones internacionales pueden desempeñar funciones sustitutas pero, si la capacidad estatal es fuerte, pueden intervenir en una modalidad de colaboración (Ramalingam y Mitchell, 2014).

Existen múltiples definiciones de **acción humanitaria** en respuesta a una crisis. Una de las más integrales y más ampliamente utilizadas es la definición de la Iniciativa de la Buena Donación Humanitaria (2003).



### **Definición: acción humanitaria**

Los objetivos de la acción humanitaria son: salvar vidas, aliviar el sufrimiento y preservar la dignidad humana tanto durante como después de una crisis y un desastre natural, así como prevenir y fortalecer los esfuerzos de preparación para dichas situaciones.

Los parámetros de lo que se denomina acción humanitaria se han expandido gradualmente. Antes se pensaba que consistía en salvar vidas únicamente, pero en la actualidad es ampliamente aceptado que incluye también salvar los medios de vida de las personas. La acción humanitaria comprende tanto asistencia como protección. Mientras que la protección ciudadana es claramente una función del Estado, actualmente se espera que toda la comunidad humanitaria – no solo instituciones como los organismos especializados de las Naciones Unidas y el

Comité Internacional de la Cruz Roja (CICR) y el Movimiento de la Cruz Roja y la Media Luna Roja más amplio – desempeñe un papel. Nuestra definición de trabajo de acción humanitaria incluye asimismo preservar la dignidad humana, de lo cual generar ingresos propios constituye una parte importante.



**Definición: protección**

Protección engloba “todas las actividades encaminadas a lograr un respeto total por los derechos de la persona, de conformidad con la letra y el espíritu de la normativa pertinente (derechos humanos, derecho humanitario y derecho de los refugiados)”. (IASC, 2011)

La acción humanitaria incluye responder a una crisis y apoyar los esfuerzos de preparación y reducción del riesgo de desastres (RRD) antes de una crisis, y de recuperación y reconstrucción después – aunque los esfuerzos de preparación y recuperación se ubican en algún punto intermedio entre actividades humanitarias e iniciativas de desarrollo a largo plazo. La importancia de trabajar en las necesidades de recuperación inmediatamente después de ocurrido un desastre natural está cobrando un reconocimiento cada vez mayor. En situaciones de conflicto y otras crisis prolongadas, la línea divisoria entre el fin de la emergencia y el inicio de la recuperación suele ser difusa. En la práctica, con frecuencia se necesitan y proveen ambos tipos de apoyo simultáneamente.



**Consejo útil**

Consideren detenidamente los alcances de su evaluación. ¿Cuál quieren que sea el foco de su intervención? ¿Les interesa más la primera fase de la respuesta a la emergencia o una fase posterior? ¿Desean incidir también en los esfuerzos de preparación? ¿Apoyar los esfuerzos de recuperación y reconstrucción? Cuanto más focalizada esté su evaluación, tanto más utilizables serán sus resultados.

La acción humanitaria debe guiarse por los siguientes principios fundamentales: humanidad, imparcialidad, neutralidad e independencia (ver el Cuadro 1). Estos son importantes para marcar una diferencia entre la acción humanitaria y otras intervenciones, por ejemplo aquellas que realizan los actores políticos y militares, y para que la acción humanitaria sea aceptada por los actores competentes sobre el terreno – por ejemplo, para garantizar el acceso a la población afectada (OCAH, 2012). Muchas organizaciones humanitarias han manifestado su adhesión a estos principios, a menudo declarando públicamente su compromiso con el Código de Conducta Relativo al Socorro en Casos de Desastre del Movimiento Internacional de la Cruz Roja y la Media Luna Roja y las Organizaciones No Gubernamentales (FICR y CICR, 1994). Los principios humanitarios son, pues, un referente clave para la evaluación de la acción humanitaria, y deben guiar asimismo la forma en que se realizan las evaluaciones (UNEG-HEIG, 2016).

**Cuadro 1:** Principios que guían la acción humanitaria (Principios Humanitarios)

Humanidad	Neutralidad	Imparcialidad	Independencia
El sufrimiento humano debe ser aliviado dondequiera que se encuentre. El propósito de la acción humanitaria es proteger la vida y la salud y asegurar el respeto por los seres humanos.	Los actores humanitarios no deben tomar partido en el caso de hostilidades ni deben participar en controversias de índole política, racial, religiosa o ideológica.	La acción humanitaria debe llevarse a cabo únicamente en respuesta a la necesidad, priorizando los casos más urgentes y absteniéndose de hacer distinciones por nacionalidad, raza, género, creencias religiosas, nivel socioeconómico o filiación política.	La acción humanitaria debe ser autónoma de los objetivos políticos, económicos, militares u otros objetivos que pudiera tener cualquier actor con respecto a las áreas donde la acción humanitaria está siendo implementada.

Fuente: OCAH (2012: 2)

Las organizaciones humanitarias se rigen asimismo por el principio de no hacer daño. En la Carta Humanitaria, este ha sido recogido en el Principio de Protección 1: “Evitar exponer a las personas a daños adicionales como resultado de nuestras acciones”. Estos daños incluyen no solo violencia y abuso de derechos sino también peligros físicos (Proyecto Esfera, 2011). Esto se explica en mayor detalle en la Sección 2, La decisión de realizar una evaluación. En la práctica común, se asume que el principio de “no hacer daño” se refiere a evitar o minimizar los efectos adversos que pudiera tener una intervención sobre la población afectada – por ejemplo, al instalar una letrina muy cerca de un pozo de agua (Christoplos y Bonino, 2016). Tal como se precisa en la Sección 2, La decisión de realizar una evaluación, este principio debe ser aplicado también a la forma en que se realiza la evaluación de una acción humanitaria (EAH).

## 1.2 Evaluación de la acción humanitaria

El Comité de Asistencia al Desarrollo de la Organización para la Cooperación y el Desarrollo Económicos (OCDE-CAD) define la evaluación como: Apreciación sistemática y objetiva de un proyecto, programa o política en curso o concluido, de su diseño, su puesta en práctica y sus resultados. El objetivo es determinar la pertinencia y el logro de los objetivos, así como la eficiencia, la eficacia, el impacto y la sostenibilidad para el desarrollo. Una evaluación deberá proporcionar información creíble y útil, que permita incorporar las enseñanzas aprendidas en el proceso de toma de decisiones de beneficiarios y donantes. La evaluación también se refiere al proceso de determinar el valor o la significación de una actividad, política o programa. (OCDE-CAD, 2002)

A partir de esta definición, nosotros definimos la EAH como sigue:



**Definición: evaluación de la acción humanitaria (EAH)**

Examen sistemático y objetivo de una intervención humanitaria para determinar el valor o la significación de una actividad, política o programa, con la intención de extraer enseñanzas que permitan mejorar las políticas y prácticas y reforzar la rendición de cuentas

Una mirada más detenida a algunos de los términos clave contenidos en esta definición revela lo siguiente:

- **Examen** – exploración o análisis para determinar el valor y la significación de una intervención.
- **Sistemático** – enfoque programado y consistente, basado en métodos creíbles.
- **Objetivo** – tomar distancia de la inmediatez de la acción humanitaria para obtener cierta perspectiva sobre ella, basando los resultados en evidencia creíble.
- **Extraer enseñanzas** para mejorar las políticas y prácticas y reforzar la rendición de cuentas, son las razones para hacer una evaluación.

La evaluación cumple dos objetivos centrales: aprendizaje y rendición de cuentas.



**Definición: aprendizaje**

Proceso mediante el cual la experiencia y la reflexión generan cambios de conducta o bien la adquisición de nuevas habilidades



**Definición: rendición de cuentas**

Rendición de cuentas es el medio a través del cual el poder se ejerce con responsabilidad. Es el proceso de tomar en cuenta las opiniones de diferentes partes interesadas y responder ante ellas, principalmente ante la población afectada por la autoridad o el poder

El grado en el que una evaluación sea verdaderamente independiente dependerá del objetivo de la evaluación (ver la [Sección 2, La decisión de realizar una evaluación](#)). Esto es más importante en el caso de evaluaciones orientadas a la rendición de cuentas. Puede ser menos alcanzable en el caso de evaluaciones orientadas al aprendizaje si las personas que realizan el aprendizaje están involucradas en la evaluación y fueron responsables de ejecutar la acción humanitaria objeto de evaluación. Sin embargo, incluso en dichos casos es deseable introducir algún nivel de objetividad en el proceso, por ejemplo incorporando al equipo a un facilitador externo o a especialistas experimentados e independientes, ya sea como parte o al mando del equipo.



**Tener presente**

Todas las evaluaciones deben esforzarse por reducir los sesgos.

**Conceptos de evaluación fundamentales**

Figura 1.1: Cadena de resultados



**Definición: insumos**

Los recursos financieros, humanos y materiales empleados en una intervención humanitaria



**Definición: productos**

Los productos, bienes y servicios que resultan de una intervención

Los productos son el resultado de los insumos recibidos y las actividades realizadas por el actor o el grupo de actores. Un producto debe ser totalmente atribuible a un actor o un grupo de actores – por ejemplo, los puntos de agua provistos por una organización de desarrollo en un campamento para desplazados internos (DI).



**Definición: efectos**

Cambios o giros intencionales o fortuitos en una situación, generados directa o indirectamente por una intervención. Pueden ser deseados (positivos) o no deseados (negativos). Pueden incluir cambios de conducta (actividades, relaciones, políticas, prácticas) en personas individuales, grupos de personas, comunidades, organizaciones, instituciones u otros actores sociales.

Un efecto es solo parcialmente atribuible al actor responsable de la intervención – por ejemplo, la forma en que se utiliza el agua de los puntos de agua recientemente instalados por una ONG (consumo doméstico, consumo animal, otras actividades generadoras de ingresos como la fabricación de ladrillos).



**Definición: impacto**

Se refiere a los efectos más amplios del programa – sociales, económicos, técnicos y ambientales – sobre personas individuales, un género determinado, grupos etarios, comunidades e instituciones. Los impactos pueden ser intencionales y fortuitos, positivos y negativos, a nivel macro (sectorial) y a nivel micro (de hogares, individual), a corto o largo plazo.

Los impactos pueden ser positivos y negativos al mismo tiempo. Por ejemplo, la distribución de ayuda alimentaria puede eximir a los hogares de tener que vender sus activos productivos – por ejemplo, sus animales – para comprar alimentos (un impacto intencional positivo), pero también puede desincentivar la producción local de alimentos (un impacto fortuito potencialmente negativo).

**Nota: El término “impacto” se utiliza a todo lo largo de esta guía para describir los efectos más amplios de la acción humanitaria, tal como se ilustra en la [Figura 18.1 de la página 357](#) (Smutylo, 2001) y acorde con la definición de impacto de la OCDE-CAD.**

**Definición: atribución**

Imputación de una relación causal entre los cambios observados (o que se espera observar) y una intervención en particular

En intervenciones humanitarias complejas, rara vez es posible atribuir un resultado a una causa específica. Una organización de ayuda alimentaria puede atribuir la reducción de la desnutrición a la distribución de alimentos, pero esta también podría deberse a otros factores: mejora de la calidad del agua, prácticas mejoradas de cuidado infantil, higiene mejorada, atención de salud, mayor oferta de saneamiento, control de vectores o incluso cambios estacionales normales. Como se demuestra en la [Figura 18.1 de la página 357](#), la atribución se torna más difícil a medida que se avanza en la cadena de resultados – y por lo tanto es más difícil atribuir los impactos que los efectos a una intervención específica.

**Definición: contribución**

Analizar la contribución en una evaluación se refiere a buscar formas creíbles de demostrar que una intervención tuvo alguna incidencia en generar los resultados. El análisis de la contribución es una especie de análisis evaluativo que reconoce que el resultado puede deberse a varias causas, incluso si estas, tomadas de manera individual, pueden no ser necesarias o suficientes para generar un impacto.

En la evaluación de la acción humanitaria suele ser mucho más fácil evaluar la contribución que la atribución.

## 1.3 El seguimiento y la evaluación en el ciclo de respuesta a una emergencia

En los proyectos y programas humanitarios la evaluación suele ser una actividad aislada, que se lleva a cabo en un momento crucial del ciclo de la respuesta humanitaria a una situación de emergencia, con la finalidad de servir de insumo tanto para el ciclo en cuestión como para respuestas futuras.<sup>1</sup> En algunos casos se pueden planificar diversas evaluaciones para distintos momentos del ciclo de una respuesta, tal como sucedió con las evaluaciones de Darfur (Broughton et al., 2006) y de la respuesta al terremoto de Haití (Grunewald et al., 2010; Hidalgo y Theodate, 2012). El seguimiento, por otro lado, debe ser permanente; es decir, debe llevarse a cabo a todo lo largo de la ejecución de proyectos y programas humanitarios. El seguimiento y la evaluación son herramientas complementarias para ayudar a determinar qué tan bien se está desarrollando una intervención (FICR, 2010:19). Como lo explica la Agencia Sueca de Cooperación Internacional para el Desarrollo (Sida),

“para que una evaluación sea factible, sin embargo, es posible que se requieran datos de seguimiento. Si una intervención no ha sido sometida a seguimiento apropiado desde un inicio, quizás no pueda ser evaluada satisfactoriamente después. Así como el seguimiento necesita de la evaluación como complemento, la evaluación necesita del respaldo del seguimiento”. (Molund y Schill, 2007: 15)



### **Definición: seguimiento**

Función sostenida que utiliza una recopilación sistemática de datos sobre indicadores señalados para proporcionar, a los administradores y a las principales partes interesadas de una intervención humanitaria en curso, indicaciones sobre el grado de avance, el logro de los objetivos y los progresos en el uso de los fondos asignados (basado en la definición de seguimiento de la OCDE-CAD, 2002)

Por lo general, los responsables de ejecutar un programa son los que deben hacerle seguimiento, para asegurar que se mantenga fiel a sus objetivos.<sup>2</sup> Por ejemplo, el seguimiento de un programa de ayuda humanitaria de emergencia probablemente se enfocaría en los insumos, es decir, en los volúmenes de ayuda alimentaria distribuidos, así como en los productos, es decir, en el número de personas que reciben ayuda alimentaria. Ocasionalmente recoge también los efectos y el impacto – por ejemplo, si los beneficiarios están vendiendo los alimentos (esto es, cómo los están utilizando – el efecto), así como el impacto en los precios de los alimentos en el mercado. En un programa de transferencias de dinero en efectivo, el seguimiento podría medir cuántas personas recibieron el dinero así como cuándo y qué cantidades. En ese caso, el seguimiento se focaliza en los insumos y en los productos. El seguimiento puede abarcar también los efectos – por ejemplo, qué hizo la gente con las transferencias de dinero en efectivo.

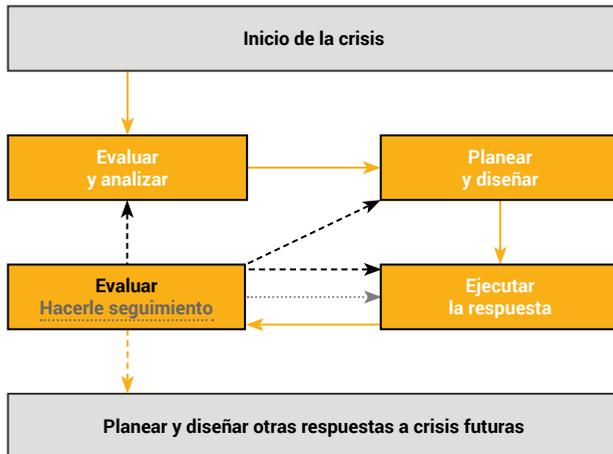
El seguimiento con frecuencia es débil en las intervenciones humanitarias, y la falta de buena información de seguimiento a menudo genera problemas para la evaluación de los programas. Las evaluaciones suelen ser realizadas por personas que no están involucradas en la ejecución de los programas, con el fin de infundirles mayor credibilidad.

La evaluación se focaliza especialmente en los efectos y el impacto: en el ejemplo de las transferencias de dinero en efectivo antes mencionado, podría examinar las consecuencias de otorgar las transferencias a mujeres, así como los impactos más amplios en el mercado y el comercio.

La [Figura 1.2](#) muestra el seguimiento y evaluación (SyE) en el ciclo de respuesta a una situación de emergencia. Como indican Hallam y Bonino (2013), los desafíos que enfrenta la EAH son un reflejo de los retos y complejidades que afronta la acción humanitaria. Algunos de los desafíos más comunes a los que debe hacer frente una EAH se describen en la [Sección 1.4, Desafíos comunes de la EAH](#).

No todas las evaluaciones de acciones humanitarias tienen que ver con la programación y el ciclo de una respuesta humanitaria, sin embargo. Las evaluaciones pueden enfocarse también en las políticas, por ejemplo (ver la [Sección 4, Tipos de evaluación](#)). Tal como se explora en la [Sección 3, Pensar desde el principio y a menudo en el uso de la evaluación](#), es importante considerar dónde encajan los resultados de una evaluación dentro de los procesos programáticos u organizativos más amplios. Para cerciorarnos de que una evaluación será útil, y será utilizada, es preciso agendar las fechas en que los resultados podrán contribuir mejor a la toma de decisiones clave.

**Figura 1.2:** El seguimiento y la evaluación en el ciclo de respuesta a una emergencia



## 1.4 Desafíos comunes de la EAH

La EAH enfrenta dos grandes grupos de desafíos: aquellos que son comunes a todas las evaluaciones, que se suelen agudizar en contextos humanitarios, y aquellos que se suscitan específicamente en la evaluación de acciones humanitarias, la misma que a menudo se lleva a cabo en entornos difíciles. Esta guía apunta a incidir en los desafíos específicos de la EAH. El listado a continuación describe algunos de los desafíos más comunes y sus posibles soluciones, con enlaces a las secciones donde aparecen de forma más detallada en la guía.<sup>3</sup>

Los evaluadores deben explicitar desde un inicio las limitaciones que enfrentan y cómo tienen previsto lidiar con ellas desde el punto de vista de la gestión de riesgos. El informe final debe indicar cómo afectaron las referidas limitaciones no solo el proceso en sí sino también a los resultados de la evaluación.

### La urgencia y el caos de las emergencias humanitarias

Una respuesta humanitaria rápida por definición se planifica a una gran velocidad, con frecuencia en condiciones extremas. Los documentos de planificación y seguimiento pueden ser escasos, los objetivos pueden ser confusos y los planes de respuesta iniciales pueden perder vigencia rápidamente, a medida que se producen cambios en el contexto o se aclara la situación. De modo que, ¿qué utilizar como punto de referencia o de partida para la evaluación?

#### Posible solución

Consideren desarrollar una cronología de la crisis y de la intervención humanitaria a ser evaluada. Entrevisten al personal de la organización para identificar los objetivos reales o implícitos y cómo podrían haberse modificado con el tiempo. Ver la [Sección 10: Métodos documentales](#).

### La inseguridad puede implicar falta de acceso en entornos de conflicto

Los problemas de seguridad dificultan o imposibilitan el acceso de los evaluadores a la población afectada, especialmente en situaciones de conflicto. ¿Cómo llevar a cabo una evaluación si no se tiene acceso a la población afectada?

#### Posible solución

Explore formas creativas de llevar a cabo la evaluación a distancia. Al presentar sus hallazgos y escribir su informe, detallen las restricciones enfrentadas. Eviten generalizar sus hallazgos, absteniéndose de hacerlos extensivos a los lugares y poblaciones que no pudieron visitar. Ver la [Sección 15: Acceso restringido](#).

## Ausencia de una línea de base

Es posible que la información haya sido destruida durante la crisis, o que haya perdido relevancia – por ejemplo, si un gran porcentaje de la población se ha visto forzada a desplazarse. En el breve plazo de una evaluación, ¿cómo recopilar toda la información que se necesita si no se cuenta con una línea de base como referencia?

### Posible solución

Entrevisten a la población afectada por la crisis y a informantes locales clave para preguntarles en qué medida ha cambiado la situación y las razones de dichos cambios. En otras palabras, recurran al recuerdo. Incluyan preguntas similares en las encuestas. Ver la [Sección 14: Involucrando a la población afectada en su evaluación](#), y el [ejemplo de buena práctica de la página 270](#).

## Alta rotación de personal

La alta rotación del personal de asistencia humanitaria, en especial personal internacional, puede plantear dificultades para los evaluadores a la hora de localizar y entrevistar informantes clave. ¿Cómo localizar a los miembros del personal que revisten importancia para el período que están evaluando?

### Posible solución

Inviten a informantes clave – ya sea exempleados o empleados ausentes – a participar en una entrevista telefónica (o por Skype) o en una encuesta en línea. Apóyense más en el personal nacional, cuya rotación suele ser menor – para lo cual posiblemente deban traducir las encuestas en línea, etc. Ver la [sección 13, Metodologías de campo](#).

## Crisis humanitarias en lugares remotos y con la infraestructura dañada

Algunas crisis humanitarias ocurren en lugares alejados y/o donde la infraestructura ha quedado dañada, lo cual dificulta el acceso. ¿Cómo llegar a dichos lugares y cómo planificar el trabajo de campo?

### Posible solución

Planifiquen el trabajo de campo cuidadosamente con alguien que esté familiarizado con la situación actual, incluyendo probables líneas de tiempo y restricciones al acceso, o consideren la posibilidad de que las visitas de campo sean programadas por la organización contraparte, conforme a criterios establecidos por el equipo de evaluación. Incorporen tiempo adicional de contingencia a su plan. Ver la [Sección 15, Restricciones al acceso](#).

## Opiniones polarizadas a consecuencia del conflicto

Los conflictos a menudo exageran las diferencias de opinión entre las personas; los acontecimientos, como de hecho la crisis misma, pueden ser interpretados de maneras radicalmente distintas. ¿Cómo realizar una evaluación “objetiva” en estos casos?

### Posible solución

Familiarícense con los puntos contenciosos del conflicto y recojan la mayor cantidad posible de opiniones diferentes, en especial de los grupos que se encuentran en orillas opuestas del conflicto, cerciorándose de que sus puntos de vista estén representados en el informe final de evaluación. Garantizar la independencia de la evaluación puede ser difícil en estas circunstancias. Si sienten que la evaluación se encuentra comprometida – por ejemplo, si les hubiera sido negado el acceso a ciertas áreas en disputa – cerciórense de dejar constancia de ello en el informe de evaluación y consideren las implicaciones de este hecho para sus hallazgos. Ver la [Sección 14: Involucrando a la población afectada en su evaluación](#).

## Pérdida de confianza a consecuencia de la politización y el trauma

En los conflictos y otras crisis humanitarias que se politizan, y donde el abuso y el maltrato y los acontecimientos traumáticos y atemorizantes son generalizados, se produce una pérdida total de confianza entre la población afectada. ¿Cómo recabar información certera y profunda de la población afectada en una evaluación corta?

### Posible solución

Diseñen metodologías de levantamiento de información y formas de involucrar a la población afectada que sean sensibles al trauma y el miedo. Consideren parejas entrevistador-entrevistado que tengan en cuenta el enfoque de género. Ver la [Sección 11, Diseños de evaluación para responder las preguntas de evaluación](#) y [Sección 13, Metodologías de campo](#).

## Los trabajadores de asistencia humanitaria operan en entornos de gran presión y estrés

Las organizaciones y trabajadores humanitarios pueden estar esforzándose por implementar sus programas en un entorno sometido a grandes presiones. Programar tiempo para pasarlo con el equipo evaluador puede no ser prioritario, especialmente en [evaluaciones en tiempo real \(ETR\)](#) al inicio de una crisis. ¿Qué metodologías son apropiadas en dichos contextos?

### Posible solución

Sean receptivos a las presiones que los trabajadores humanitarios pueden estar enfrentando y busquen formas más bien “informales” de conectar con ellos. Por ejemplo, los ejercicios breves de reflexión y aprendizaje pueden sacar a la luz revelaciones interesantes si están dotados de una moderación adecuada. Ver la [Sección 2: La decisión de realizar una evaluación](#). Pregunten al personal qué desea saber acerca de la intervención y en qué áreas desean modificar el enfoque de su organización para acometer una tarea.

## Presiones de tiempo sobre la población afectada

Las personas pueden tener poco tiempo para participar en una evaluación, dado que su supervivencia es prioritaria. ¿Cómo lograr que la población afectada participe en la evaluación de una manera apropiada?

### Possible solución

Eviten utilizar métodos largos, como discusiones de grupos focales, consultando más bien a la población afectada mientras se encuentra haciendo cola – esperando la distribución de la ayuda o un medio de transporte – y recurriendo a la observación directa. Tengan en cuenta las presiones de tiempo al diseñar la evaluación: por ejemplo, ¿es posible retrasar la evaluación hasta que haya pasado el pico más alto de la respuesta, una vez que la población se haya calmado y esté sometida a menos presiones? Ver la [Sección 14: Involucrando a la población afectada en su evaluación](#).

## Falta de responsabilidades claramente delimitadas entre los actores humanitarios

La falta de responsabilidades claramente delimitadas entre los actores humanitarios puede entorpecer la rendición de cuentas y la atribución del impacto, especialmente en crisis de gran envergadura, que concitan la participación de muchos actores humanitarios. ¿Cómo atribuir los resultados y el impacto?

### Possible solución

Enfóquense en la contribución antes que en la atribución. Consideren llevar a cabo una evaluación conjunta para explorar el impacto combinado de varias organizaciones, o de la respuesta humanitaria internacional en su totalidad, en vez de intentar aislar el impacto de una sola organización por medios artificiales. Ver la [Sección 4: Tipos de evaluación](#).

## Términos de referencia demasiado ambiciosos y recursos limitados

Los TdR pueden suponer una carga de trabajo poco realista, por ejemplo si se espera que una tarea que normalmente requeriría el equivalente de un año de trabajo a cargo de dos personas sea ejecutada con un presupuesto para el equivalente de cuatro meses de trabajo por una sola persona (ver la [Sección 7](#), que trata sobre los TdR). ¿Cómo manejar las expectativas y producir una evaluación creíble?

### Possible solución

Utilicen el informe inicial para pulir la tarea de la evaluación de modo tal que pueda ser realizada en los plazos y con los presupuestos asignados. Se trata de una manera provechosa de manejar las expectativas desde una etapa temprana. Ver la [Sección 8: Fase inicial](#).

## Restricciones éticas a los enfoques experimentales

La hipótesis, “si las personas afectadas por la sequía no hubieran recibido ayuda alimentaria, muchas de ellas habrían muerto”, se denomina contrafactual. Por razones éticas, desde luego, la validación de esta hipótesis absteniéndose de prestar ayuda a algunas personas afectadas no constituye una opción. En las [Secciones 2 y 14](#) se analizan consideraciones éticas, mientras que a lo largo de toda la guía se presentan otras consideraciones éticas específicamente ligadas a la EAH, bajo el ícono .

### Posible solución

Comparen distintas formas de asistencia (por ejemplo, donaciones en efectivo, entrega de vales, ayuda alimentaria). Aprovechen cualquier “experimento” que pudiera haber ocurrido naturalmente, por ejemplo el que un grupo de personas no recibiera ninguna asistencia debido a que se encontraba en un lugar carente de acceso. Ver la [Sección 11: Diseños de evaluación](#) para responder las preguntas de evaluación.

## Evaluando la protección

Los desafíos en este rubro incluyen la falta de claridad respecto de qué significa protección y cómo medir lo que no es cuantificable – especialmente ¿qué no pasó? Por ejemplo, violencia o abuso y maltrato como resultado de las acciones de las organizaciones humanitarias. Ver la [Guía de Acompañamiento sobre Protección](#) (Christoplos y Bonino, 2016) al respecto.

A pesar de estos grandes retos, la EAH con frecuencia debe ser acometida a una gran velocidad. En consecuencia, la EAH requiere de competencias de evaluación que trasciendan aquellas que normalmente se requieren para otras evaluaciones (ver la [Sección 9, Planificación y gestión de su evaluación](#)), a fin de asegurar que el rigor y la credibilidad no se vean comprometidos (UNICEF, 2013: 3-4).

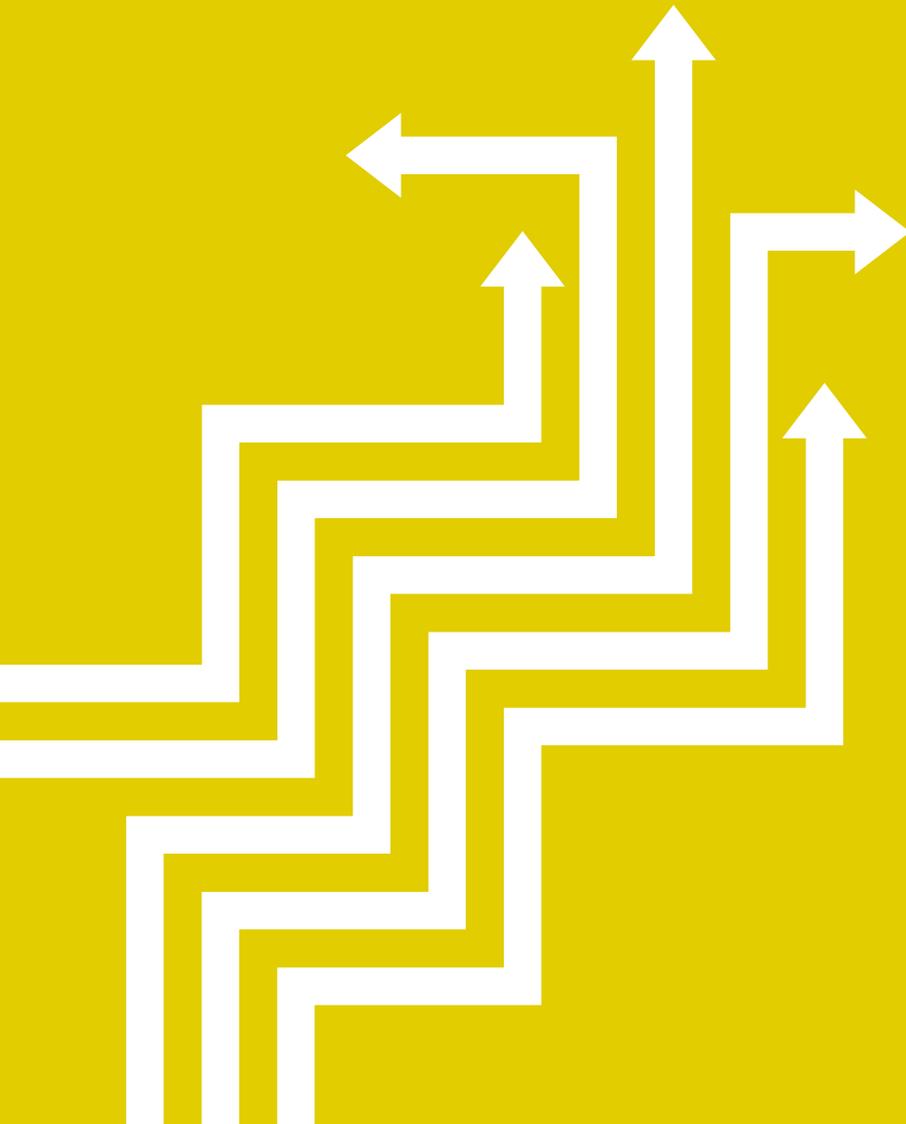
## Notas finales

1. Esta sección se basa en: Morra Imas y Rist (2009) y Hallam (2011).
2. Existen formas externas de seguimiento, tales como formularios de calificación de la ayuda recibida por las personas beneficiarias (Banco Mundial, 2004), pero estas generalmente se utilizan en contextos de desarrollo antes que en humanitarios. El seguimiento a cargo de terceros está cobrando cada vez mayor importancia en emergencias complejas como la de Siria (ver la investigación de SAVE sobre el tema en: <http://www.save.gppi.net/home/>).
3. Esta sección ha sido adaptada de Beck (2006).

## Notas

## Notas

# La decisión de realizar una evaluación



# 2 / La decisión de realizar una evaluación

En ocasiones la contratación de una evaluación responde a una obligación contractual contraída con una institución donante, pero en otros casos es posible que esta decisión deba ser sopesada a la luz de otras opciones. La presente sección presenta algunas opciones, las mismas que dependerán del objetivo de la actividad evaluativa. Si usted es el coordinador de un programa o de una evaluación, esta sección le ayudará a decidir si una evaluación en efecto constituye la mejor opción y, de ser así, cuándo hacerla. También le brindará orientación sobre cómo realizar un “análisis de evaluabilidad”.

## 2.1 ¿Cuándo es apropiada una evaluación para la rendición de cuentas y el aprendizaje?

### Rendición de cuentas

La evaluación es uno de varios procesos destinados a satisfacer los requisitos de rendición de cuentas de una organización, es decir, su obligación de responder por su trabajo frente a otros, por ejemplo ante su propio directorio, las instituciones donantes o la población afectada. Ustedes deberán decidir si se trata de la opción más apropiada y eficaz en función de los costos.

El Cuadro 2.1 ofrece un resumen de distintos tipos de rendición de cuentas y sugiere otras opciones, además de la evaluación. Por ejemplo, si el foco de la atención en términos de rendición de cuentas y cumplimiento es de índole financiera – frente a las entidades donantes, la alta gerencia o el directorio de una organización – en ese caso una opción más apropiada puede ser una auditoría. Una auditoría evalúa la solvencia y el cumplimiento de los requisitos de una organización y es por lo tanto distinta a una evaluación, que emite un juicio de valor (FICR, 2010). La rendición de cuentas frente a la población afectada (RCPA) está recibiendo mucha mayor atención actualmente, tal como se describe en la Sección 14, Involucrando a la población afectada en su evaluación. Pero esta se logra principalmente mediante un proceso permanente de retroalimentación y consulta a todo lo largo de la ejecución de una intervención (Darcy et al., 2013). Las evaluaciones pueden medir la RCPA y deberían asimismo apoyar este tipo de rendición de cuentas (ver HAP, 2010), pero considerar esta opción justo después de haberse embarcado en una EAH es demasiado tarde.

La rendición de cuentas a menudo tiene dimensiones externas e internas. La gerencia de una organización puede tener que responder por el uso de los recursos frente a las instituciones donantes (externa) así como frente a su directorio (interna), ya que el uso indebido de los recursos puede representar un peligro para la supervivencia de la organización. Del mismo modo, la rendición de cuentas frente a la población afectada puede obedecer a obligaciones contraídas por la gerencia y el directorio.

De acuerdo a Sandison (2006):

El impacto de la evaluación se ve potenciado, si no posibilitado, por el hecho de formar parte de un menú más amplio de enfoques encaminados a mejorar el desempeño. El seguimiento, por ejemplo, sigue siendo un pariente pobre de la evaluación y pasará algún tiempo antes de que reciba la misma atención de parte de las instancias de decisión. La evaluación, en cualquiera de sus formas, es solo un componente de la rendición de cuentas (p. 140).



#### Preguntas clave

Si la rendición de cuentas es el objetivo principal de su evaluación, pregúntense lo siguiente:

- ¿Qué tipo de rendición de cuentas es el que más les preocupa, y rendición de cuentas frente a quién (con referencia al [Cuadro 2.1](#))?
- ¿Es la evaluación la opción más idónea para cumplir con ello?
- ¿La evaluación debería articularse con algún otro proceso de rendición de cuentas? (Ver el [ejemplo de buena práctica](#) a continuación.)



#### Ejemplos de buenas prácticas: Articulando mecanismos de rendición de cuentas: auditoría y evaluación

Para la evaluación de dos de sus programas nacionales, la Organización para la Alimentación y la Agricultura (FAO) había efectuado una auditoría poco antes de dichas evaluaciones independientes (Somalia en 2012 y Sudán del Sur en 2015). Esta secuencia surtió efecto en el sentido de que los aspectos de eficacia en función de los costos y cumplimiento, especialmente en relación con la eficacia y la eficiencia de los procesos y procedimientos de gestión, habían sido abordados por la auditoría antes de la evaluación. Esto liberó al equipo de evaluación para enfocarse en los aspectos de programación estratégica, pudiendo aprovechar asimismo algunos de los hallazgos de la auditoría.

**Cuadro 2.1:** La evaluación y otros procesos de rendición de cuentas

Tipo de rendición de cuentas	¿Es apropiada una evaluación?	Otros procesos de rendición de cuentas que considerar
Rendición de cuentas frente a la población afectada	Sí, si la evaluación se diseña adecuadamente, con total participación de la población afectada	El diálogo y la consulta permanentes con la población afectada, así como el uso de mecanismos de retroalimentación a lo largo de toda la vida del proyecto o programa, pueden constituir formas más efectivas de rendir cuentas frente a la población afectada que una evaluación.
Rendición de cuentas estratégica (por ejemplo, con referencia al mandato y los objetivos de la organización)	Sí	Revisión estratégica
Rendición de cuentas sobre la gestión (por ejemplo, sobre el uso de los recursos en una organización)	Sí	Gestión del desarrollo y otras herramientas de gestión Auditoría de desarrollo
Rendición de cuentas financiera (por ejemplo, frente a las entidades donantes) y de cumplimiento financiero (por ejemplo, frente a la alta gerencia y al directorio)	Sí, especialmente si la eficacia en función de los costos y la eficiencia constituyen la principal preocupación	Auditoría (puede ser una opción más apropiada si el control y el cumplimiento de las normas y procedimientos financieros constituyen la principal preocupación)
Rendición de cuentas contractual (por ejemplo, para cumplir obligaciones contractuales)	Sí, especialmente si existe la obligación contractual de realizar una evaluación	Auditoría; otros procesos establecidos en el convenio o contrato
Rendición de cuentas relacional (por ejemplo, frente a otras organizaciones involucradas en una operación)	Sí, si está incluido en los <u>términos de referencia</u> de la evaluación	Examen institucional
Rendición de cuentas legal (por ejemplo, con relación a la legislación nacional o internacional)	No	Auditoría de cumplimiento de leyes (por ejemplo, de la legislación laboral)

En los hechos, las instituciones donantes son una de las partes interesadas más poderosas. Gran parte de la actividad evaluativa tiende a obedecer por lo tanto a la necesidad de rendir cuentas a las entidades donantes (Brown y Donini, 2014). Las evaluaciones orientadas a la rendición de cuentas pueden utilizarse también para responder ante la alta gerencia o el directorio de una organización. La focalización recae en los resultados, a menudo examinando cuán adecuadamente se utilizaron los recursos para cumplir con la responsabilidad y el mandato de la organización. Por esta razón, las evaluaciones orientadas a la rendición de cuentas suelen efectuarse hacia la mitad o hacia el final de un programa, cuando ya debe haber resultados sobre los cuales informar.

### Aprendizaje

Las evaluaciones orientadas al aprendizaje están pensadas para facilitar el aprendizaje grupal, individual y/o institucional. Las evaluaciones orientadas al aprendizaje representan una valiosa oportunidad para aprender a partir de casos reales. Pueden ser muy efectivas para examinar las cosas que han y que no han funcionado bien y cómo mejorarlas. La focalización con frecuencia recae en programas que se encuentran en la fase inicial y de ejecución, a fin de identificar lo que está funcionando bien y lo que puede ser mejorado. Pueden llevarse a cabo en cualquier momento del ciclo de un programa. En el sector humanitario en particular, donde tanto los recursos para aprendizaje como los antecedentes de aprendizaje son escasos (ALNAP, Red de Evaluación del Desarrollo del CAD/OCDE y UNEG, 2010; Patrick, 2011; Knox-Clarke y Darcy, 2014), las evaluaciones pueden ser sumamente útiles para generar conocimiento e iniciar procesos de aprendizaje institucional.

La evaluación no es, sin embargo, la única forma de promover el aprendizaje institucional, ni tampoco necesariamente la más eficaz en función de los costos. En el Cuadro 2.2 se presentan otros procesos de aprendizaje que considerar. Algunos de ellos pueden ser integrados en una evaluación orientada al aprendizaje – por ejemplo, los exámenes retrospectivos (para mayores detalles sobre estas y otras metodologías orientadas al aprendizaje, ver la [Sección 13.7](#)).

Cuadro 2.2: Procesos de aprendizaje

Metodología de aprendizaje	Breve descripción	Cómo puede utilizarse en la evaluación
<u>Examen retrospectivo</u>	Proceso facilitado por un moderador para que los participantes en un programa reflexionen acerca de lo ocurrido, los éxitos y desafíos y el aprendizaje	Este proceso podría ser facilitado por los evaluadores como parte de una evaluación orientada al aprendizaje, y el aprendizaje resultante podría ser incluido en el informe de evaluación (USAID, 2006; Ramalingam, 2006: 64).
<u>Contar historias con uso de metáforas</u>	Proceso facilitado por un moderador y en el que los participantes "cuentan la historia" de lo que pasó y extraen aprendizajes	La metodología de contar historias utilizando metáforas puede emplearse con el personal, tal como se describe en el <u>ejemplo de buena práctica de la página 260</u> , para incentivar la reflexión y el aprendizaje compartido. La plataforma en línea <i>Better Evaluation</i> ofrece algunas referencias útiles (Better Evaluation, 2014).
<u>Metodología del "cambio más significativo" (CMS)</u>	Enfoque participativo que consiste en recolectar historias de "cambios significativos" a nivel de campo. Las historias más importantes son seleccionadas por paneles de partes interesadas como una forma de recoger información sobre los efectos y el impacto de un proyecto (Davies y Dart, 2005).	La metodología del CMS puede ser utilizada con las personas afectadas por una crisis para identificar los cambios que estas atribuyen a la intervención (tal como sucedió con la evaluación de la FICR de la respuesta al terremoto en Perú de 2007, Martínez, 2009).
Investigación apreciativa	Enfoque que prioriza las soluciones antes que los problemas, involucrando a partes interesadas clave en un ejercicio de reflexión acerca de las cosas que funcionaron y qué aprendizajes pueden extraerse de ello	Aunque la investigación apreciativa es poco frecuente en la EAH, Preskill y Catsambas (2006) realizan un análisis detallado de esta metodología y ofrecen ejemplos de su utilización en evaluaciones de distintos sectores.

Algunas organizaciones, como UNICEF y Tearfund, emplean procesos de aprendizaje independientes en las respuestas pequeñas y medianas a situaciones de emergencia (por ejemplo, exámenes retrospectivos) y tratan de alentar ejercicios de reflexión y aprendizaje. Sin embargo, en las respuestas a emergencias de gran magnitud UNICEF utiliza la evaluación, incluyendo evaluaciones en tiempo real o ETR (ver la Sección 4, Tipos de evaluación).

Una evaluación no generará por sí sola aprendizajes institucionales, aunque quienes participan en y se comprometen con una evaluación pueden extraer aprendizajes personales. Una organización con una cultura de aprendizaje (Cuadro 2.3) será más propensa a beneficiarse de una evaluación. La publicación

de ALNAP, “Usando evaluaciones para un cambio real” (Hallam y Bonino, 2013), identifica los siguientes factores como fundamentales para que las organizaciones humanitarias realmente utilicen las evaluaciones: un liderazgo apoyador, una cultura institucional de evaluación, estructuras organizativas que promuevan la evaluación y la capacidad para garantizar los recursos adecuados, humanos y financieros.

**Cuadro 2.3:** La evaluación y los tres componentes básicos de una organización que aprende

Componente básico	Características distintivas	Implicaciones para la evaluación
Un entorno de aprendizaje favorable	El personal se siente lo suficientemente seguro como para discrepar con otros, hacer preguntas ingenuas, reconocer sus errores y representar puntos de vista minoritarios. Reconoce el valor de las ideas contradictorias y está dispuesto a asumir riesgos y a explorar lo desconocido.	Utilicen metodologías de evaluación que incentiven la franqueza, tales como: investigación apreciativa, la metodología del cambio más significativo y exámenes retrospectivos.
Procesos de aprendizaje concretos	Existen procesos formales para generar, recopilar, interpretar y difundir información.	Consideren pedir a algunos miembros del personal que se integren al equipo de evaluación. Establezcan un proceso formal para involucrar al personal en el análisis de los hallazgos y el desarrollo de las conclusiones.
Un liderazgo que refuerza el aprendizaje	La alta gerencia demuestra que está dispuesta a considerar puntos de vista alternativos; emite señales de que reconoce la importancia de dedicar tiempo a la identificación de problemas, la transferencia de conocimiento y la reflexión; y realiza activamente preguntas y escucha las respuestas.	Otorguen un decidido apoyo gerencial a la evaluación y asignen el tiempo necesario para ello. Creen un espacio de retroalimentación y análisis de los resultados con la alta gerencia.

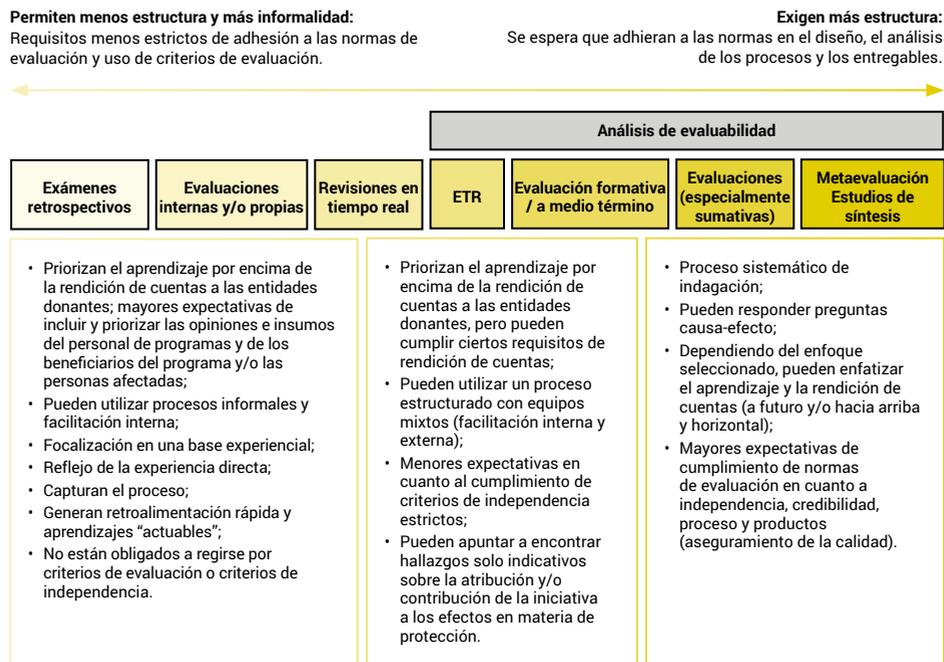
Fuente: Adaptado de Garvin et al. (2008)

## 2.2 ¿Es la evaluación la herramienta adecuada para la tarea?

¿Están seguros de que deben hacer una evaluación? La evaluación cuesta dinero que de otra manera podría ser invertido en la intervención humanitaria, para prevenir muertes o aliviar el sufrimiento. El dinero para evaluación *estará* bien gastado si conduce a mejoras en la acción humanitaria, pero esto solo ocurrirá si se toman medidas concretas en respuesta a los hallazgos. Por eso, es importante seleccionar la herramienta adecuada.

La [Figura 2.1](#) ilustra el espectro de opciones, desde ejercicios de aprendizaje reflexivo más informales a la izquierda hasta enfoques evaluativos más formales y estructurados, requeridos especialmente para evaluaciones orientadas a la rendición de cuentas, a la derecha. Una vez que hayan definido el objetivo general de la evaluación, probablemente deseen considerar en qué parte del espectro podría ubicarse su actividad evaluativa.

**Figura 2.1 Espectro que va desde ejercicios de aprendizaje reflexivo hasta opciones evaluativas más formales**



Fuente: Bonino (2016), ampliado de Scharbatke-Church (2011: 7)

## 2.3 Estableciendo un balance entre aprendizaje y rendición de cuentas

Tal como sugiere la Figura 2.1, muchas evaluaciones sirven en la práctica a un objetivo dual de rendición de cuentas y aprendizaje. Como señala Irene Guijt: “No puede haber rendición de cuentas si no hay aprendizaje. Y, para aprender, tienes que saber si te estás desempeñando a la altura de las expectativas” (2010: 277), aunque lograr ambos objetivos puede ser difícil en los hechos. No obstante lo anterior, es importante determinar cuál es el objetivo principal o predominante, ya que esto ayudará a determinar la selección del diseño de evaluación.

Al planificar y diseñar una evaluación, estas son las preguntas clave que deben hacerse:



### Preguntas clave

- ¿Qué objetivo es más importante – rendir cuentas o aprender? ¿En qué lugar del espectro entre rendición de cuentas y aprendizaje se ubica? No debe ubicarse en el medio, o correrán el riesgo de perder perspectiva.
- Si el objetivo principal es la rendición de cuentas, ¿rendición de cuentas frente a quién y para qué?
- Si el objetivo principal es el aprendizaje, ¿aprendizaje de quién? ¿Y cómo se supone que tenga lugar?

### Características de las evaluaciones orientadas a la rendición de cuentas y al aprendizaje

Una vez establecido el objetivo principal de su evaluación, podrán diseñar el proceso de evaluación así como las metodologías para llevarlo a cabo. Cada objetivo de evaluación tiene características específicas propias. Por ejemplo, una evaluación orientada a la rendición de cuentas pondrá un mayor énfasis en la objetividad y la autonomía y adoptará un estilo más inquisitivo, buscando atribuir la responsabilidad tanto por los éxitos como por los fracasos. Pero no es seguro que conduzca al aprendizaje, si hace que las personas que necesitan aprender se pongan a la defensiva. Para aprender se requiere un entorno psicológico seguro (Edmondson, 2014), donde reconocer las dificultades y admitir los errores sea aceptable. Una evaluación orientada al aprendizaje será más propensa por lo tanto a utilizar un estilo más facilitador, alentando a participar y reflexionar a las personas involucradas en la ejecución. Estas diferencias explican la posible dificultad de combinar el aprendizaje y la rendición de cuentas en el objetivo de una evaluación. Ver el Cuadro 2.4 para una descripción de cómo el enfoque y el estilo de una evaluación pueden variar entre evaluaciones orientadas a la rendición de cuentas o al aprendizaje.

**Cuadro 2.4:** Evaluaciones centradas en la rendición de cuentas y el aprendizaje

Elemento de la evaluación	Evaluación orientada a la rendición de cuentas	Evaluación orientada al aprendizaje
Términos de referencia	Basados en insumos de partes interesadas externas y de los directores de programas	Deberían ser definidos por las personas directamente involucradas en el programa que desean o necesitan aprender
Composición del equipo	Normalmente personal externo independiente	Personal interno, quizás con un facilitador o jefe externo, o bien una mezcla de personal interno y externo
Énfasis en el enfoque	Metodologías de recolección y análisis de datos (más objetivas)	Proceso enfocado en reflexionar y extraer conclusiones (más participativo)
Metodologías	Una combinación de metodologías cuantitativas y cualitativas que generarán evidencia robusta	Metodologías participativas que involucran a los que van a aprender
Coordinación	Los responsables de rendición de cuentas	Los responsables de gestión del conocimiento y aprendizaje
Estilo de gestión	Más directivo	Más facilitador
Divulgación del informe	Normalmente de dominio público	Puede constreñirse a la organización, para incentivar una participación franca y abierta

## 2.4 Decidiendo qué evaluar y cuándo

Para cerciorarse de que una evaluación sea útil, y que sea utilizada, es importante programar las fechas en que los resultados pueden contribuir de manera óptima a la toma de decisiones cruciales. Como explican Hallam y Bonino (2013), un factor muy importante en la demanda de una evaluación son las fechas de la evaluación misma así como de la presentación de sus resultados. Es importante que los responsables de contratar una evaluación tengan en cuenta este factor. Muchos potenciales usuarios de evaluaciones se quejan de que estas a menudo llegan demasiado tarde como para ser de alguna utilidad para la toma de decisiones (Weiss, 1983b; 1990).

Un informe sobre las investigaciones y evaluaciones del DFID dice lo siguiente: La crítica más común de las personas entrevistadas a las evaluaciones fue el cronograma: si bien este marca ritmos importantes en las oficinas nacionales, en general se consideró que se extendían demasiado en el tiempo como para ser relevantes para los equipos de políticas, y que no se presta la debida atención para su articulación con los ciclos de políticas y ventanas de oportunidad. (Jones y Mendizábal, 2010: 11)

Resulta contraproducente encargar evaluaciones para cada proyecto o programa si estas tenderán a sobrecargar a la organización y si se cuenta con sistemas de seguimiento eficaces. Las evaluaciones estratégicamente seleccionadas tienen mayores probabilidades de satisfacer necesidades de conocimiento esenciales de manera oportuna.

Preguntas clave que hay que tomar en cuenta para seleccionar las evaluaciones estratégicamente:



#### **Preguntas clave**

- ¿Cómo puede añadir valor el proceso de evaluación a la organización en general?
- ¿Cuántas evaluaciones tiene capacidad de absorber la organización?

La unidad de evaluación de la cooperación sueca Sida ha probado con éxito un enfoque estratégico para la selección de las evaluaciones humanitarias y de desarrollo. Este enfoque podría ser aplicado de manera similar solo a las EAH.



#### **Ejemplo de buena práctica: Selección estratégica de evaluaciones**

El ciclo de planificación de evaluaciones de Sida comienza con una discusión de la unidad de evaluación con todas las unidades operativas sobre lo que desearían saber y cómo una evaluación podría contribuir a ello. Se confecciona una relación de alrededor de 100 ideas preliminares de evaluaciones, de las cuales la unidad escoge 15, que se suman a las cerca de 80 evaluaciones que se realizan todos los años a nivel operativo (Sida, 2011).

El Consejo Noruego para Refugiados (NRC) prioriza las evaluaciones “en función de su relevancia – evaluando las cosas apropiadas en el momento apropiado”, lo cual se identifica mediante un plan anual de evaluación (NRC, 2015).

Realizar una evaluación puede ser especialmente apropiado en los siguientes casos:

- Para programas cuyos efectos se desconocen o son controvertidos;
- Para intervenciones costosas de gran magnitud;
- Para iniciativas piloto, o para validar una idea para un programa nuevo;
- En áreas específicas en las que la organización tiene un interés estratégico;
- En áreas específicas sobre las cuales las partes interesadas desean una evaluación.

Realizar una evaluación resulta inapropiado en los siguientes casos:

- Cuando no tiene visos de aportar conocimiento nuevo;
- Cuando los problemas de seguridad o la falta de información socavarían su credibilidad.

Weiss (1998) añade algunos casos en los que realizar una evaluación representaría un ejercicio fútil por la dinámica institucional interna. Algunos ejemplos: si el gestor de la evaluación (o alguien similar) limita radicalmente lo que la evaluación está autorizada a examinar, excluyendo algunos temas importantes, o si no hay suficiente dinero o personal calificado para llevarla a cabo.

Programar una evaluación en una etapa inicial de la ejecución de un programa humanitario permite establecer los sistemas de seguimiento necesarios desde un inicio, y puede informar e influir también en el diseño de la evaluación final.

Algunas organizaciones establecen detonantes para ciertos tipos de evaluaciones, especialmente los organismos de las Naciones Unidas. Por ejemplo, el umbral de la FAO para realizar una evaluación es un gasto de 10 millones de dólares o más, mientras que para UNICEF lo son las emergencias humanitarias de envergadura. Estos detonantes tienen que ver principalmente con la rendición de cuentas, cuando se han invertido grandes cantidades de recursos, y en cierta medida también con un posible riesgo para la reputación de la organización.

Muchos de estos aspectos pueden ser recogidos en una política de evaluación. Una política de evaluación puede orientar asimismo muchas de las decisiones que se describen en esta sección con respecto al objetivo de una evaluación, si es conveniente o no realizar una evaluación y cuándo, así como algunas de las consideraciones éticas y de aseguramiento de la calidad que deben ser tomadas en cuenta. Si están mandando hacer una evaluación, la política institucional de evaluación debe figurar en los términos de referencia de la evaluación. Los evaluadores deben familiarizarse con la política de evaluación institucional desde un principio, ya que puede contener orientaciones y requisitos sobre el diseño y la metodología. Ver el ejemplo de buena práctica a continuación, referido a la política de evaluación de NRC, sucinta pero exhaustiva.



### Ejemplo de buena práctica: La política de evaluación de NRC

La política de evaluación de NRC explica por qué, cómo y cuándo utiliza evaluaciones el NRC. Además, establece normas de conducta y describe la contribución de las evaluaciones al aprendizaje y la rendición de cuentas institucionales (pág. 3). Se trata de un documento de 11 páginas dirigido al personal responsable de encargar, planificar y utilizar evaluaciones. Establece cuatro principios básicos que deben apuntalar el trabajo de evaluación: adecuación, flexibilidad, transparencia y participación, en reconocimiento de los desafíos específicos de evaluar el trabajo de NRC con personas desplazadas internas, que pueden haber experimentado hechos traumáticos, y otros desafíos más generales de la EAH, como acceso restringido y contextos rápidamente cambiantes. La política establece requisitos mínimos de evaluación, por ejemplo la obligación de todas las oficinas nacionales de encargar al menos una evaluación externa cada tres años. Ofrece orientaciones para asegurar la utilidad de las evaluaciones y establece los criterios de calidad que deben ser tenidos en cuenta.

Fuente: NRC (2015)

### Cómo realizar un análisis de evaluabilidad<sup>1</sup>

Es conveniente realizar un análisis de evaluabilidad antes de tomar la decisión de emprender una evaluación, a fin de considerar algunos de los factores mencionados arriba. El análisis de evaluabilidad es un proceso descriptivo y analítico destinado a producir una argumentación razonada para proceder a realizar una evaluación (Schmidt, Scanlon y Bell, 1979), y también puede emplearse para sugerir opciones para maximizar la viabilidad y la utilidad de una evaluación a ser contratada más adelante (Davies, 2013). El propósito principal del análisis no es solo decidir si la evaluación puede ser acometida o no, sino también garantizar que durante la ejecución se tomen medidas para crear las condiciones necesarias para facilitar el proceso de evaluación (UNEG 2011: 17). Si se concluye que la evaluación debe proceder, el trabajo realizado como parte del análisis de evaluabilidad puede informar directamente los términos de referencia de la evaluación.

Aunque infrecuente en la EAH, el análisis de evaluabilidad podría realizarse al poco tiempo de iniciada una intervención humanitaria, con miras a mejorar el diseño del proyecto y por lo tanto su evaluabilidad subsiguiente, por ejemplo creando sistemas de seguimiento fuertes y monitoreando indicadores definidos. Más común en la EAH son los análisis de evaluabilidad que apuntan a establecer si los plazos para la evaluación son correctos, o si esta debería posponerse hasta una etapa posterior o hasta que el contexto sea más oportuno, por ejemplo, de mayor seguridad.

Rick Davies (2013) ha realizado una síntesis de los documentos de orientación disponibles sobre análisis de evaluabilidad, elaborando un desglose (ver el Cuadro 2.5) de los principales temas cubiertos y los pasos comúnmente sugeridos.

**Cuadro 2.5:** Descripción genérica de análisis de evaluabilidad

Principales temas y/o pasos	Pasos comúnmente sugeridos
Definir los límites del proyecto	<ul style="list-style-type: none"> <li>• Definir el período de tiempo, la extensión geográfica y las partes interesadas</li> <li>• Establecer los productos esperados del análisis de evaluabilidad</li> </ul>
Identificar los recursos disponibles	<ul style="list-style-type: none"> <li>• Identificar la documentación disponible</li> <li>• Identificar las partes interesadas</li> </ul>
Identificar y revisar la documentación	<ul style="list-style-type: none"> <li>• La lógica/teoría del cambio/cadena de resultados del programa así como la claridad, plausibilidad y apropiación involucradas</li> <li>• Los sistemas de información</li> <li>• Disponibilidad, pertinencia y calidad de la información; capacidad de los sistemas y del personal para entregar lo que se necesita</li> <li>• Examinar la ejecución, contrastándola con la planificación</li> </ul>
Involucrar a las partes interesadas	<ul style="list-style-type: none"> <li>• Identificar la percepción de las partes interesadas acerca del objetivo, el diseño y la ejecución del programa, incluyendo las áreas de acuerdo y desacuerdo</li> <li>• Identificar sus expectativas de una evaluación: objetivos específicos, proceso y uso</li> <li>• Esclarecer y suplir los vacíos encontrados en la revisión de documentos</li> </ul>
Desarrollar conclusiones y formular recomendaciones	<p>Las conclusiones y recomendaciones del análisis de evaluabilidad deben cubrir:</p> <ul style="list-style-type: none"> <li>• Mejoras a la lógica del proyecto</li> <li>• Sistemas y desarrollo de capacidades de MyE (si se requiere/según se requiera)</li> <li>• Preguntas de evaluación de interés prioritario para las partes interesadas</li> <li>• Posibles diseños de evaluación solamente si se requiere/según se requiera *</li> </ul>
Comunicar los hallazgos y conclusiones a las partes interesadas	

\*Davies advierte a los responsables de encargar o ejecutar un análisis de evaluabilidad sobre la tentación de aventurarse en territorio reservado al diseño de evaluación, empezando a realizar sugerencias de posibles diseños. Fuente: Davies, 2013: 16

Para EAH, el análisis de evaluabilidad podría cubrir cuatro áreas fundamentales:

1. **Nivel general de aspiración y tipo de preguntas** que las partes interesadas de la evaluación y del programa quisieran que respondiera la evaluación.
2. **Diseño del programa y lógica de la intervención** – Se trata de un área particularmente importante para las evaluaciones de efectos y de impacto que utilizan diseños basados en teorías para desentrañar la causalidad, así como para diseños de metodología mixta y enfoques centrados en efectos que examinan la contribución a los resultados de intervenciones multisectoriales o interinstitucionales (por ejemplo, mapeo de efectos; cosecha de efectos; el enfoque de mapeo de efectos del programa RAPID).

3. **Disponibilidad de información** o viabilidad de generar información con los recursos asignados, de modo tal que la evaluación pueda responder las preguntas de evaluación seleccionadas.
4. **Idoneidad del contexto** para llevar a cabo la evaluación, por ejemplo, en términos de acceso, logística y seguridad, y también si la oficina local tiene la capacidad para albergar al equipo de evaluación, así como el “clima” institucional y el apoyo del liderazgo necesarios.

La cuarta área – idoneidad del contexto (ver la Sección 15, Acceso restringido) – es la razón más común para decidir no seguir adelante con una evaluación o postergar su ejecución. Los análisis de evaluabilidad pueden ser útiles cuando no se sabe con certeza si será posible o no llevar a cabo una evaluación creíble dadas las circunstancias – un ejemplo de ello es el análisis de evaluabilidad de Somalia (Cosgrave, 2010). En ocasiones, el análisis de evaluabilidad se combina con revisiones iniciales del programa, como la revisión programática y el estudio de evaluabilidad del programa de educación en emergencias y transición poscrisis de UNICEF (Barakat et al., 2010) y la revisión preparatoria para la evaluación del Clúster Global de Educación (Reid et al., 2010).

En la práctica, el análisis de evaluabilidad normalmente se utiliza en programas humanitarios a gran escala y para evaluaciones multiinstitucionales. En la mayoría de las EAH, especialmente a pequeña escala, un “examen de evaluabilidad rápido” realizado por el equipo de evaluación durante la fase inicial suele ser suficiente. Dicho examen puede recomendar mejoras a la cobertura y a la focalización de la evaluación. Una vez que el equipo de evaluación ha sido designado, suele ser muy tarde para evaluar la idoneidad del contexto y considerar suspender o postergar la evaluación. Si están encargando una EAH a pequeña escala y tienen reparos acerca de la idoneidad del contexto, deberán considerar un “examen de evaluabilidad rápido” antes de designar al equipo de evaluación.



#### **Ejemplo de buena práctica: “Examen de evaluabilidad rápido”**

Durante la fase inicial de una evaluación conjunta comisionada por la Red de Comunicación con Comunidades Afectadas por el Desastre (CDAC) para evaluar las iniciativas de “Comunicación con las Comunidades” (CwC) en respuesta al terremoto de 2015 en Nepal, el equipo de evaluación llegó a la conclusión de que el objetivo superior y los objetivos específicos establecidos en los términos de referencia requerían de un estudio de investigación para validar las hipótesis, antes que una evaluación. Esto se conversó con el Grupo de Coordinación de la evaluación conjunta durante la fase inicial y se decidió por unanimidad que, en efecto, debía considerarse un estudio de investigación, y no una evaluación.<sup>2</sup>

Un análisis de evaluabilidad o un “examen de evaluabilidad rápido” pueden garantizar por consiguiente que la EAH haya sido adaptada al contexto – por ejemplo, al tipo de crisis y al tipo de respuesta bajo evaluación, así como al contexto institucional, por ejemplo, a la existencia o no de una cultura de aprendizaje a nivel institucional y un compromiso a nivel del liderazgo.

## 2.5 Ética y EAH

Varios organismos de evaluación han producido guías sobre la ética en la evaluación. Por ejemplo, la guía ética Principios Rectores para Evaluadores, de la Asociación Americana de Evaluación, enfatiza los siguientes principios: indagación sistemática, competencia, integridad, honestidad, respeto por las personas y responsabilidad por el bienestar general y público.<sup>3</sup> Los lineamientos éticos del Grupo de Evaluación de las Naciones Unidas (UNEG, 2008) se focalizan en la intencionalidad de la evaluación, las obligaciones de los evaluadores, las obligaciones frente a los participantes y el proceso y el producto de evaluación.

Algunas organizaciones humanitarias han desarrollado lineamientos éticos específicos para la EAH. Por ejemplo, la Federación Internacional de Sociedades de la Cruz Roja y la Media Luna Roja (FICR) se refiere al principio de “no hacer daño”, el mismo que se aborda en mayor detalle en la [página 56](#), y alude también al Principio 5 del Código de Conducta Relativo al Socorro en Casos de Desastre del Movimiento Internacional de la Cruz Roja y la Media Luna Roja y las Organizaciones no Gubernamentales, que dice lo siguiente: “Nos empeñaremos en respetar la cultura, las estructuras y las costumbres de las comunidades y los países en donde ejecutemos actividades.” (FICR, 2010)

Los principios éticos de la evaluación del NRC incluyen:

- No hacer daño a los informantes y a las partes interesadas;
- Al entrevistar a niños y niñas y sobrevivientes de la violencia de género, ceñirse a las normas internacionales; y
- Enfocarse en las preguntas de evaluación y evitar afligir a la gente (por ejemplo, no pedirles que revivan hechos traumáticos). (NRC, 2015)

Al realizar un análisis de evaluabilidad, puede ser importante tener en cuenta estos lineamientos éticos. Por ejemplo, ¿es posible ejecutar el trabajo de campo sin poner en riesgo la integridad de los informantes y las partes interesadas? [El recuadro de la página 57](#) incide en este aspecto en mayor detalle. Los aspectos éticos de involucrar a la población afectada en una EAH se tratan con mayor detenimiento en la [Sección 14, Involucrando a la población afectada en su evaluación](#).

## La evaluación debe respetar el principio de “no hacer daño”

Las EAH deben observar una serie de principios éticos, pero deben sujetarse en todos los casos al principio de no hacer daño, al igual que las operaciones humanitarias bajo evaluación. Esto rige especialmente para las evaluaciones que se efectúan en situaciones de conflicto e inseguridad. El punto de partida debe ser considerar de qué manera intervenir en el proceso de evaluación podría perjudicar a las personas que participan en ella o son consultadas.<sup>4</sup>

No importa si una evaluación examina la respuesta a un conflicto o a un desastre natural, los evaluadores deberán tener en cuenta en todos los casos cómo podría exacerbar las tensiones al:

- Generar expectativas de que participar en la evaluación (por ejemplo, durante la etapa de levantamiento de información) hará que se les proporcione una mayor asistencia, lo cual podría generar sentimientos de frustración;
- Desatar discusiones acaloradas, por ejemplo entre diferentes grupos de la población afectada, durante el proceso de levantamiento de información, al incidir en temas que agudizan las tensiones y divisiones;
- Generar la impresión de estar reuniendo información para una de las partes del conflicto;
- Presentar involuntariamente un análisis sesgado al representar inadecuadamente las opiniones de las diferentes partes interesadas.

Para evitar dichos riesgos y acometer la evaluación con una actitud “respetuosa del conflicto”, los gestores de la evaluación y los evaluadores deberán considerar desde un inicio:

- Evaluar si alguna etapa del proceso de evaluación podría contribuir a avivar las tensiones (especialmente durante las etapas de levantamiento de información y difusión del informe de evaluación);
- En situaciones de conflicto, realizar un nuevo análisis del conflicto (o actualizar un análisis ya existente), para proporcionar información a la planificación y al diseño de la EAH (ver Chigas y Goldwyn, 2012);
- Revisar los planes de evaluación a la luz de dicho análisis, asegurando que no contribuyan a avivar las tensiones (y, de ser posible y si se encuentra dentro del mandato de la organización, intentando minimizarlas).

Fuente: adaptado de Bonino (2016)



## En profundidad: ¿Qué significa “no hacer daño”?

La expresión “no hacer daño” tiene significados diferentes en los ámbitos de la acción humanitaria y la sensibilidad a los conflictos. “No hacer daño” se deriva del principio médico según el cual un profesional de la salud no debe causar ningún daño al paciente. Ha sido adoptado y adaptado en otros campos.

**Desde una perspectiva humanitaria**, “no hacer daño” es un término ampliamente difundido, pero a menudo no está claramente definido. En el manual del Proyecto Esfera, dicho término se recoge en el Principio de Protección 1: “Evitar exponer a las personas a daños adicionales como resultado de nuestras acciones”. Dichos daños adicionales incluyen la violencia y el abuso de los derechos, así como las amenazas físicas. En la práctica común, en ocasiones se ha usado “no hacer daño” para transmitir la idea de evitar o minimizar los efectos negativos de una intervención. Por ejemplo, la instalación de una letrina demasiado cerca de un pozo en ocasiones ha sido calificada de “daño”.

**Desde una perspectiva de sensibilidad a los conflictos**, No Hacer Daño es un marco específico de siete pasos que puede ser utilizado para evaluar la sensibilidad a los conflictos de una intervención. Fue desarrollado por Collaborative for Development Action (actualmente CDA) y es la “herramienta” más ampliamente utilizada para evaluar la sensibilidad frente a los conflictos.

Sensibilidad a los conflictos quiere decir cerciorarse de que una intervención no contribuya por error a un conflicto y que, dentro de lo posible, contribuya a establecer la paz (dentro de los alcances del mandato de la organización). Según esta definición, “no hacer daño” tiene que ver con los riesgos asociados a un conflicto, entre ellos muchos riesgos de protección, ya que ambos se encuentran estrechamente ligados.

Cabe mencionar que existen muchas otras herramientas para el abordaje de la sensibilidad a los conflictos, más allá del modelo de No Hacer Daño, y que existen muchas prácticas y análisis que inciden en la sensibilidad a los conflictos más en general.

Fuente: Christoplos y Bonino (2016)

## Notas finales

1. Basado en gran medida en Christoplos y Bonino (2016)
2. Ver el portal web de la red CDAC: [www.cdacnetwork.org](http://www.cdacnetwork.org).
3. Ver: [www.eval.org/p/cm/ld/fid=51](http://www.eval.org/p/cm/ld/fid=51)
4. Estos aspectos también son abordados en la guía de obligaciones éticas de UNEG para los responsables de iniciar, coordinar y ejecutar evaluaciones (UNEG, 2008). Dichas obligaciones incluyen: respeto a la dignidad y la diversidad; derechos humanos; confidencialidad; y no hacer daño.

## Notas

## Notas

# Iniciando la evaluación



# 3/ Pensar desde el principio y a menudo en el uso de la evaluación

El objetivo de esta sección es ayudar a los responsables de encargar y coordinar evaluaciones a considerar desde un principio qué necesitan hacer para cerciorarse de que los hallazgos y recomendaciones de la evaluación sean utilizados.

El contenido de la sección se basa en gran medida en el trabajo previo de ALNAP sobre el uso de las evaluaciones, el mismo que tiene por objeto mejorar la forma en que las organizaciones humanitarias hacen suyos los hallazgos de una EAH y los utilizan efectivamente (Sandison 2006; Hallam, 2011; Hallam y Bonino, 2013).

## 3.1 Qué significa estar enfocados en el uso

Una evaluación enfocada en el uso se realiza teniendo en mente a los usuarios primarios previstos y contemplando usos puntuales, declarados y prácticos.



Sandison, con base en Patton (1997) y Carlsson et al. (1999), ofrece la siguiente definición de uso: “Una evaluación ha sido usada si los usuarios, que tienen la intención y el potencial para tomar acción, han prestado atención seria y activa a sus hallazgos, identificando usos que tengan sentido para sus propios intereses y necesidades”. (2006: 100-101)

Es fundamental identificar a los usuarios previstos desde un inicio, para ayudarles a decidir qué desean lograr con la evaluación. La participación de los usuarios en la decisión relativa al objetivo principal, los objetivos específicos y los alcances de una EAH debe guiar la elección del diseño y las metodologías. La focalización en el uso y la colaboración con los usuarios primarios deben seguir guiando el proceso de evaluación, desde la fase de planificación hasta la ejecución. Esta forma de pensar y trabajar requiere de compromiso y tiempo, pero garantiza que la EAH contribuya a enriquecer el conocimiento de sus usuarios y a generar cambios y mejoras en los hechos (Hallam y Bonino, 2013).



### Tener presente

Es importante recordar que la evaluación cuesta dinero, un dinero que de otra manera podría ser invertido en la intervención humanitaria, para prevenir muertes o aliviar el sufrimiento. El dinero para evaluación estará bien gastado si conduce a mejoras en la acción humanitaria, pero la evaluación solo podrá mejorar la acción humanitaria si se toman en serio los hallazgos y se hacen suyas las recomendaciones.

## 3.2 Cómo se utilizan las evaluaciones de acción humanitaria

Las EAH se utilizan en general de dos maneras: para fines de rendición de cuentas (también llamados sumativos) y para fines de aprendizaje, normalmente enfocados en un proyecto específico (en cuyo caso se denominan formativos) pero en ocasiones focalizados en un conocimiento más general, a nivel sistémico (desarrollistas). La mayoría de las EAH contienen una combinación de elementos, aunque uno de ellos posiblemente predomine (ver el Cuadro 3.1).

Cuadro 3.1: Objetivos de una evaluación

Uso	Preguntas	Ejemplos
<b>Sumativo</b> Juzgar la valía o el valor de un programa (por ejemplo, si cumple con su deber de rendir cuentas a las partes interesadas o de justificar las decisiones financieras)	¿El programa satisface las necesidades? ¿Tiene valía? ¿Cuáles son sus efectos?	La evaluación conjunta de la asistencia humanitaria y los esfuerzos de reconstrucción en Afganistán en el período 2001-2005 (DANIDA, 2005)
<b>Formativo</b> Potenciar el aprendizaje (por ejemplo, con la intención de mejorar un programa)	¿Qué cosas funcionan y no funcionan? ¿Cuáles son las fortalezas y debilidades actuales?	La <i>Revisión de aprendizaje institucional de la respuesta internacional de Caritas a la emergencia del tsunami</i> (Otto et al., 2006) facilitó el aprendizaje, privilegiando la apertura y la participación de las partes interesadas más importantes.  Las evaluaciones en tiempo real suelen tener un objetivo de aprendizaje primario (Saavedra, 2013).
<b>Desarrollista</b> Contribuir a generar conceptos, ideas y enfoques nuevos, ya sea dentro de una organización o a nivel de todo el sector	¿El programa tiene en cuenta los hechos y limitaciones del mundo real? ¿Cuáles son los patrones comunes que atraviesan todos los programas?	La evaluación conjunta de Ruanda (Borton et al., 1996), que introdujo nuevas ideas con respecto a la rendición de cuentas en las organizaciones humanitarias y precipitó importantes innovaciones de política tales como las normas mínimas del Proyecto Esfera.

El uso de las EAH puede subdividirse aún más:

- **Uso instrumental:** Implementación directa de los hallazgos y recomendaciones por las instancias de decisión, por ejemplo para realizar modificaciones a un programa en curso o enfocarse más en la práctica. La mayor parte de las EAH han sido diseñadas para este tipo de utilización, estén centradas en la rendición de cuentas o en el aprendizaje.
- **Uso conceptual:** Los resultados y conclusiones de la evaluación son asimilados por la organización en forma de ideas y conceptos nuevos, tal como se describe bajo la categoría “desarrollista” en el Cuadro 3.1. Este uso es menos común y a menudo se utiliza de manera acumulativa. Los cambios conceptuales pueden no ser atribuibles a una sola evaluación, pero las síntesis evaluativas pueden ser particularmente útiles para incentivar un cambio conceptual.
- **Uso del proceso o del aprendizaje:** El involucramiento y la participación en la evaluación propiamente dicha pueden generar aprendizajes personales y propiciar cambios de conducta.
- **Uso de legitimación:** La evaluación legitima una decisión ya tomada o una percepción ya existente en una organización o una oficina nacional, al suministrar evidencia independiente y objetiva que puede ser utilizada para justificar acciones posteriores (ver Sandison, 2006; Hallam y Bonino, 2013: 18).

En ocasiones las evaluaciones se utilizan indebidamente por varias razones:

- **Uso ritualista:** La evaluación se lleva a cabo con fines meramente simbólicos, por ejemplo para cumplir con una obligación contractual frente a la organización donante, pero con escaso o ningún compromiso institucional para hacerla suya.
- **Uso indebido:** Los hallazgos son ocultados, tergiversados o distorsionados con fines políticos propios o personales.
- **No uso:** Los hallazgos son ignorados porque los usuarios encuentran poco o ningún valor en ellos (debido, por ejemplo, a que las recomendaciones están mal formuladas), o bien no saben de ellos, o el contexto ha variado radicalmente. El no uso puede deberse a la falta de convencimiento de la dirección, el diseño deficiente de la evaluación, la falta de respuesta de la evaluación a sus propias preguntas o su incapacidad para presentar evidencia convincente con respaldo de sus conclusiones (Sandison, 2006).

## 3.3 Identificar, entender e involucrar a los usuarios previstos de la evaluación

Esta subsección ofrece consejos útiles a los responsables de encargar y coordinar evaluaciones, concentrándose en las fases de diseño y planificación para asegurar que la evaluación se enfoque en el uso.

Los tres pasos principales son los siguientes:

1. Identificar a las partes interesadas de la evaluación y, dentro de ese grupo, a los usuarios primarios previstos.
2. Consultar a los usuarios previstos cuáles son sus necesidades de información, asegurando al mismo tiempo que los alcances de la evaluación se mantengan realistas y manejables.
3. Buscar formas de involucrar a los usuarios previstos a lo largo de toda la evaluación.

Estos pasos también pueden ser útiles para el equipo de evaluación. Si no han sido tomados para la fecha de nombramiento del equipo de evaluación, el equipo deberá facilitar su realización antes de seguir adelante con el proceso de evaluación. Por ejemplo, durante la fase inicial, el equipo de evaluación debe facilitar un ejercicio de mapeo de las partes interesadas con el gestor de la evaluación (ver información más adelante) y consultar con los usuarios primarios previstos para averiguar qué desean de la evaluación.

### Identificar a las partes interesadas

El primer paso consiste en identificar a las partes interesadas y a los usuarios de la evaluación. Una parte interesada es una persona u organización que tiene un interés o algo en juego en la evaluación debido a que puede verse afectada por las decisiones que surjan de sus resultados o recomendaciones.

No todas las partes interesadas utilizarán necesariamente la evaluación final, sin embargo. Un buen ejemplo de ello es la población afectada. Son partes interesadas porque la evaluación les ofrece la oportunidad de comunicar sus puntos de vista, y puede resultar en mejoras al programa, pero es poco probable que utilicen sus resultados finales.

### Llegar a los usuarios previstos

Los usuarios previstos son los destinatarios más importantes de una EAH. Las cuatro preguntas a continuación les ayudarán a identificar a los usuarios primarios previstos y también pueden usarse en un ejercicio visual participativo o un análisis del poder y de los intereses de las partes interesadas.

**1.****¿Quiénes son sus partes interesadas para los fines de la evaluación?**

Recuerden que se trata de las partes interesadas de la evaluación, y no de la intervención humanitaria objeto de evaluación.

Las partes interesadas pueden comprender un espectro que va desde la unidad de evaluación que encarga la evaluación, pasando por el personal ejecutor del programa y las autoridades locales, hasta la población afectada.

**2.****Partes interesadas con un interés directo en la evaluación**

Las partes interesadas que tienen un interés directo en la evaluación deben ser plenamente involucradas en la misma.

Podría tratarse de las entidades financieras del programa y del personal ejecutor del programa.

**2.****Partes interesadas con un interés indirecto**

Se trata de aquellas partes interesadas que deberían ser influenciadas por la evaluación o que son consultadas (por ejemplo, el personal que trabaja para otras organizaciones presentes en la zona) y que deberían ser consultadas, por ejemplo la población afectada.

**3.****¿Qué partes interesadas con un interés directo en la evaluación son sus usuarios previstos?**

En la mayoría de los casos, si la idea es que aprendan de los hallazgos de la evaluación, los usuarios previstos son el personal ejecutor del programa. En una evaluación orientada a la rendición de cuentas, los usuarios previstos pueden ser también las entidades financieras del programa.

**4.****¿Quiénes son sus usuarios primarios previstos?**

Esta pregunta cobra especial importancia si existen varias partes interesadas primarias con intereses diversos y antagónicos. Si no identifican y priorizan claramente entre los usuarios previstos, los diferentes objetivos antagónicos pueden devenir inmanejables, y es posible que no logren alcanzar adecuadamente ninguno de ellos.

Las entidades financieras pueden tener un interés especial en la efectividad y en el coste-eficiencia del programa, mientras que el personal de programas puede querer aprender sobre lo que funcionó y no funcionó bien y cómo mejorar su desempeño, y la alta gerencia puede interesarse por las implicaciones de política más amplias de los hallazgos de la evaluación.



**Consejo útil:** Consideren de qué forma pueden ser influenciadas o incentivadas las partes interesadas para que los usuarios previstos (por ejemplo, la alta gerencia) realmente hagan uso de los hallazgos de la evaluación.

En su informe de evaluación debe constar claramente quiénes son sus usuarios primarios previstos. Esto puede hacerse de forma muy sencilla o más elaborada, tal como se puede apreciar en los dos ejemplos a continuación.

- “Los usuarios primarios de la evaluación serán las partes interesadas del Ministerio de Relaciones Exteriores y las contrapartes ejecutoras de Danida, mientras que los usuarios secundarios probablemente serán el Parlamento y el público en general.” (Mowjee, Fleming y Toft, 2015: 16).
- “Esta evaluación tiene múltiples partes interesadas. En primer lugar está la oficina de Concern Worldwide en la República Democrática del Congo [RDC], en cuyo caso la evaluación servirá para informar los programas actuales y futuros de la institución en Masisi. En segundo lugar está la sede central de Concern Worldwide, en cuyo caso contribuirá al aprendizaje institucional desde la RDC y su potencial aplicación en otros contextos. El tercer grupo de partes interesadas está formado por las entidades donantes de Concern en la RDC, es decir, la OFDA [la Oficina de Asistencia para Desastres en el Extranjero de USAID], que financia la evaluación; ECHO [el Departamento de Ayuda Humanitaria y Protección Civil de la Comisión Europea]; y Irish Aid, que financia también el trabajo humanitario de Concern en Kivu del Norte. Por último, la evaluación puede ser de utilidad para otros actores humanitarios interesados en las enseñanzas prácticas y en la evidencia de intervenciones basadas en la entrega de dinero en efectivo.” (Bailey, 2013: 10).

Para evaluaciones más grandes y complejas, posiblemente deseen considerar un análisis de partes interesadas utilizando una “matriz del poder y los intereses” clásica, tal como se muestra en la [Figura 3.1](#). Dicha matriz fue desarrollada durante la fase inicial de la evaluación humanitaria interinstitucional de la respuesta a la crisis de República Centroafricana durante el período 2013-2015. Ayudó al equipo de evaluación a entender a las diferentes partes interesadas, involucrarlas sin sobrecargarlas y promover una participación estructurada de los usuarios. Después de una minuciosa evaluación de las partes interesadas según su respectivo poder e intereses (ver el Anexo 2 del informe inicial), la matriz propuso a los coordinadores de evaluación, en este caso OCHA (la Oficina de Coordinación de los Asuntos Humanitarios de la ONU), una estrategia para promover la participación de los usuarios. Por ejemplo, los evaluadores se enfocaron en “monitorear” al gobierno y a la población afectada de República Centroafricana (RCA), asumiendo un interés bajo en la evaluación (pese a lo cual podrían beneficiarse de la oportunidad de aportar puntos de vista, aprender de la respuesta, medir el éxito de la intervención e informar estrategias futuras) y un poder también bajo para facilitar el proceso (pese a lo cual los evaluadores necesitaban de su participación en el levantamiento de información y la aceptación formal del proceso) (Lawday, 2015; comunicación personal con Lawday, noviembre de 2015).

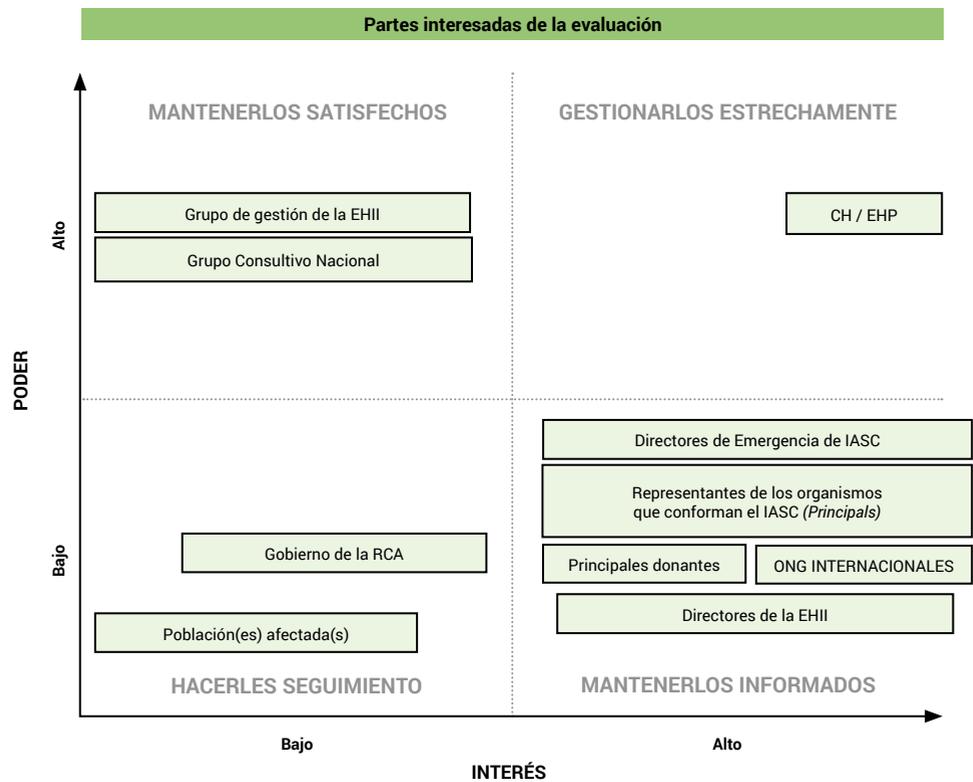
## Mapeo de partes interesadas

Los diferentes grupos de partes interesadas pueden ser mapeados mediante una serie de círculos concéntricos en los cuales se distingue a las partes interesadas primarias, que supuestamente deben estar totalmente involucradas en la evaluación, y las partes interesadas con un interés indirecto, que pueden cumplir un rol distinto. Las partes interesadas que tienen un interés indirecto se subdividen en aquellas que deberían ser influenciadas por la evaluación además de consultadas y aquellas que deberían ser solo consultadas.

Una forma de realizar esta clase de mapeo es anotar a todas las partes interesadas en notas adhesivas, dibujar círculos en un papelógrafo y luego ubicar cada nota en el círculo indicado, en función de los intereses y la pretendida participación de la parte interesada en la evaluación. Este puede ser un ejercicio participativo rápido, y puede usarse también en evaluaciones pequeñas.



Figura 3.1 Análisis de las partes interesadas de la evaluación



Fuente: Lawday (2015). Para mayores detalles sobre esta herramienta, ver Start y Hovland (2004: 26).

### ¿Cuáles son las necesidades de información y evidencia de los usuarios?

Una vez que hayan identificado a los usuarios primarios previstos de su evaluación, pregúntales: ¿Qué necesitan saber para poder decidir mejor qué hacer y cómo hacerlo? Generalmente formulada por el responsable de encargar o coordinar la evaluación, esta pregunta contribuye a esclarecer y verbalizar los objetivos de la evaluación.

Si les preguntan a las partes interesadas qué les gustaría saber, los alcances de la evaluación pueden multiplicarse fácilmente, llevando a un listado de temas inmanejable. Anímelos a concentrarse en el uso, haciendo hincapié en el “conocimiento requerido para tomar acción”, al formular a los usuarios primarios previstos las siguientes preguntas:

- ¿Qué necesitan saber para marcar una diferencia?
- ¿Cómo utilizarán los hallazgos de la evaluación?

Las respuestas deberían acercarlos al objetivo y los alcances de la evaluación, y servirles de base para decidir la forma más idónea de presentar la información y la evidencia recopiladas por medio de la evaluación para facilitar su utilización. Por ejemplo, ¿cuáles deberían ser el formato y/o la extensión del informe? ¿Qué otras formas de difusión funcionarían con los usuarios previstos? ¿Se avecinan reuniones clave o foros donde se tomarán decisiones importantes en los cuales se podrían presentar los hallazgos de manera inmediatamente pertinente? Para un análisis más detallado de las distintas opciones, ver la [Sección 17: Informes y comunicación de los resultados de evaluación con un enfoque de utilización](#).

### Involucrando a los usuarios primarios previstos

Las personas son más proclives a utilizar una evaluación si entienden el proceso y los hallazgos y se apropian de ellos, lo que a su vez es más probable que suceda si han sido activamente involucradas lo largo de toda la evaluación (Patton, 1997). Esto significa que se requieren interacciones repetidas entre los usuarios previstos, de un lado, y los gestores de la evaluación y evaluadores, del otro. Esto requiere una inversión de tiempo y recursos.



#### Pregunta clave

Examinen su proceso de evaluación. ¿Qué pasos o actividades puntuales podrían ser más participativos o incluir a los usuarios?

He aquí algunas ideas para involucrar a los usuarios primarios previstos, que se abordan en mayor detalle más adelante:

- Involucrar a los potenciales usuarios en la elaboración de los [términos de referencia](#).
- Considerar incorporar a algunos usuarios al [equipo de evaluación](#). (Las organizaciones tienen políticas variadas sobre esto, y algunas prefieren que los usuarios previstos participen en grupos de referencia externos, antes que en el equipo de evaluación.)
- Constituir un [grupo de referencia](#) para la evaluación que incluya a algunos usuarios primarios previstos.
- Organizar un taller para que los usuarios primarios previstos y el equipo de evaluación diseñen de manera conjunta la evaluación. (Ver el [ejemplo de buena práctica en la página 72](#).)
- [Compartir el borrador del informe](#) con todas las personas que hayan sido entrevistadas y darles la oportunidad de comentarlo.
- Involucrar a las partes interesadas en el [análisis de la información recogida](#).
- Organizar un taller para [presentar los hallazgos de la evaluación](#).
- Pedir a los usuarios que participen en la [elaboración de las recomendaciones](#).
- Pedir a los usuarios que diseñen una [estrategia de difusión](#).

Es importante asimismo considerar actividades que abarquen todo el proceso de evaluación, por ejemplo:

- Cerciorarse de que el equipo de evaluación mantenga una comunicación regular y transparente con los usuarios previstos a todo lo largo del proceso de evaluación. Esto ayudará a mantener la expectativa y evitar sorpresas desagradables. Mantener informados a los usuarios previstos de la evaluación con respecto a los avances de la evaluación y las modificaciones a los alcances o metodologías, por ejemplo, facilitará el uso de los hallazgos.
- Involucrar a los usuarios primarios previstos en las decisiones importantes, por ejemplo cualquier cambio en la focalización de la evaluación durante la fase inicial si los términos de referencia originales resultan demasiado ambiciosos para los recursos disponibles. La integración de usuarios en el grupo de referencia de la evaluación puede ser útil para ello.

Las evaluaciones pueden generar ansiedad o resistencia entre quienes están siendo evaluados. Estos sentimientos pueden agudizarse en situaciones de crisis humanitarias de perfil alto, que concitan atención mediática internacional. En estos casos es más importante aún construir un sentido de apropiación. Esto puede equilibrarse con la necesidad de mantener la objetividad o la independencia de la evaluación, si los evaluadores se rigen por normas y lineamientos profesionales. Objetividad no significa desentenderse de las necesidades de los usuarios. Los evaluadores tienen que tener las habilidades interpersonales necesarias para establecer relaciones constructivas con usuarios clave y crear un clima de confianza. La comunicación frecuente a lo largo de toda la evaluación ayuda. Generar un compromiso con la evaluación a menudo implica promover la apertura al cambio.



#### **Ejemplo de buena práctica: Involucrando a partes interesadas clave a todo lo largo de la evaluación**

La evaluación del Proyecto de Desarrollo de Capacidades en Situaciones de Emergencia sobre la respuesta al terremoto de Yogyakarta (Wilson et al., 2007) estuvo fuertemente enfocada en el uso. Algunas de sus características centrales incluyeron:

- diálogo desde un inicio con los miembros del comité de coordinación en el país, para averiguar qué deseaban de la evaluación;
- diálogo entre el jefe del equipo y el comité de coordinación antes de la llegada del jefe del equipo al país;
- conversaciones entre el equipo de evaluación y el comité de coordinación en torno a las metodologías de evaluación – por ejemplo, sobre la interpretación más apropiada de los criterios del CAD.



### **Ejemplo de buena práctica: Ayudando a partes interesadas clave a definir sus objetivos**

Un ejemplo de colaboración entre las partes interesadas con el fin de determinar los objetivos de una evaluación es el estudio de aprendizaje conjunto sobre reconstrucción de viviendas tras el terremoto de Yogyakarta en 2006. Participaron en el estudio la Cruz Roja, Medair, Caritas Suiza, Solidar, Heks y Solidaridad Suiza.

Antes de la evaluación, los usuarios previstos, especialmente el personal de la sede central de las organizaciones, las personas encargadas de realizar la evaluación, ebaix y la entidad donante, Solidaridad Suiza, asistieron a un taller facilitado por terceras personas para desarrollar e identificar los temas de aprendizaje centrales. Se emplearon métodos participativos, tales como la técnica del “espacio abierto”, para que los participantes pudieran seleccionar temas de aprendizaje relevantes para su trabajo presente y futuro en programas humanitarios de reconstrucción, lo cual permitió definir las áreas de investigación. Los participantes aportaron ejemplos de proyectos para el trabajo de campo (Otto, 2014).

## **3.4 El rol de la población afectada en la fase de planificación**

Como mencionáramos anteriormente, suele haber opciones para que las organizaciones rindan cuentas a la población afectada que resultan más apropiadas que una evaluación (ver la [Sección 14, Involucrando a la población afectada en su evaluación](#)), y los ejemplos de consulta con la población afectada en torno a la focalización de la evaluación o a las preguntas que es preciso hacer son todavía escasos. [El ejemplo de buena práctica de la página 73](#) es un caso interesante pero infrecuente en que esto se hizo para una evaluación conjunta del impacto humanitario (ECIH).



### **Ejemplo de buena práctica: Consulta con la población afectada sobre las posibilidades de una ECIH**

Para averiguar si existía apoyo para la ECIH y explorar posibles enfoques, se llevaron a cabo consultas entre febrero y noviembre de 2010. Se consultó con la población afectada de 15 comunidades de Bangladesh, Haití y Sudán, así como con actores humanitarios y representantes gubernamentales en cada país. “Se trata quizás del intento más sistemático de consultar con la población afectada durante la fase de diseño de un ejercicio de evaluación de grandes proporciones” (Beck, 2011: iv). Algunos miembros de la población afectada manifestaron su deseo de participar activamente en el proceso de evaluación y de ser informados sobre los resultados de la ECIH, para intervenir en la validación de los hallazgos. Algunos se mostraron escépticos de que la evaluación pudiera hacer alguna diferencia: “Como señaló una comunidad de Sudán del Sur, les preocupaba que, pese a haber participado en varias evaluaciones, no había habido ningún cambio en los errores que habían identificado” (Ibid: iv).

Fuente: Beck (2011)

Un proceso de diálogo y consulta efectivo entre las organizaciones y la población afectada a todo lo largo de la fase de ejecución permitirá a los responsables de planear y diseñar la evaluación formarse una idea clara de los tipos de temas que deberían ser explorados. Por ejemplo, si la población afectada ha manifestado preocupación por los aspectos de protección, en ese caso puede ser apropiado incluir una pregunta de evaluación acerca de cuán eficaz ha sido el manejo de la protección por el programa. Ver la [Sección 14](#) para mayores detalles sobre cómo involucrar a la población afectada en su evaluación.

## **3.5 Factores que afectan el uso de una evaluación**

“El cambio no es un proceso racional, sino que se produce por medio de una compleja variedad de interacciones a través de una organización. Como resultado, aprovechar todo el potencial de la EAH como fuerza de cambio requiere prestar especial atención a toda una serie de cuestiones relacionadas entre sí.” (Hallam y Bonino, 2013: 18-19) La presente subsección presenta dos modelos para ayudarles a analizar los factores que influyen en el uso de una evaluación. El primero es una adaptación del modelo “contexto, evidencia y vínculos” diseñado por el programa RAPID del ODI para evaluar las conexiones entre investigaciones y políticas (Crewe y Young, 2002) y puede utilizarse para analizar detalladamente el contexto más amplio, así como las formas más idóneas de promover el uso de su evaluación dentro de ese contexto. El segundo es el marco de ALNAP sobre áreas de capacidad de evaluación (Hallam y Bonino, 2013), que se enfoca en el uso de evaluaciones a nivel institucional.

## El modelo “contexto, evidencia, vínculos”

**Calidad de la evaluación** – La calidad de todo el proceso de evaluación, y no solo del producto final, es un factor decisivo para determinar su utilidad y las probabilidades de que sea utilizada. Este es el factor sobre el cual el equipo de evaluación tiene el mayor control, y tiene un sustancial impacto sobre la credibilidad. Como concluyó Sandison, el cuadro de utilización que emerge [de los estudios sobre el tema] es complejo y a menudo inesperado. Una de las pocas certezas es que el cómo y el por qué se lleva a cabo una evaluación afecta significativamente las probabilidades de que esta sea utilizada (2006: 91).

**Contexto institucional** – Este factor engloba temas más amplios como la cultura institucional de aprendizaje y gestión del conocimiento y la estructura institucional en términos de proximidad de las unidades de evaluación a las instancias de decisión. Las evaluaciones tienen mucho mayores probabilidades de ser utilizadas en una organización que busca activamente información sobre el desempeño con el fin de mejorar la gestión y ejecución del programa. Esto puede ser fuertemente influenciado por el hecho de que líderes clave apoyen o no el aprendizaje o atribuyan un valor bajo a la toma de decisiones basada en evidencia y tiendan a estar a la defensiva frente a una evaluación.

**Relaciones y nexos** – Este factor incluye la relación entre las personas que realizan la evaluación y las usuarias de la misma. El factor personal – por ejemplo, la presencia de un paladín de la evaluación, o varios, entre las partes interesadas más importantes con las cuales es vital que los evaluadores construyan y mantengan una relación – puede ser crucial para determinar si una evaluación se utiliza o no. El cargo formal y la autoridad efectiva de dichos paladines pueden importar menos que su entusiasmo e interés.

**Factores externos** – Estos pueden incluir preguntas del público en general o de los medios, por ejemplo, en una respuesta humanitaria de perfil alto, acerca del uso que se le dio a las ingentes sumas de dinero recaudadas. Este fue el caso de la respuesta al tsunami en el Océano Índico en 2006, por ejemplo. “Factores externos” también puede referirse al contexto de políticas humanitarias más amplio y si este es abierto o no al cambio. Algunos observadores creen que el sistema humanitario es ampliamente receptivo a los cambios bajo presión cuando los factores de presión son fuertes (Buchanan-Smith, 2005: 98), como sucedió con la evaluación interinstitucional de Ruanda (ver el [ejemplo de buena práctica en la página 75](#)). Por otra parte, el informe de síntesis de la Coalición de Evaluación del Tsunami (TEC, por sus siglas en inglés) concluyó que la falta de presiones externas a favor de un cambio fue una razón importante por la que hubo tan pocas mejoras al desempeño en los 10 años que precedieron a la respuesta al tsunami (Telford y Cosgrave, 2006).

Los factores externos se encuentran representados en un gráfico de tres círculos en la [Figura 3.2](#), en que las circunstancias externas se encuentran en el área sombreada en verde claro que rodea los círculos. Al planificar su evaluación, tómense el tiempo para considerar cómo cada uno de estos círculos podría facilitar o entorpecer el uso de los hallazgos de la evaluación y qué pueden hacer para eliminar cualesquier obstáculos y promover la utilización.

Figura 3.2 Factores que afectan el uso de una EAH



Nota: Para mayores detalles sobre este modelo, ver Start y Hovland (2004: 6)



**Ejemplo de buena práctica: La evaluación conjunta de la asistencia a la emergencia de Ruanda**

La evaluación conjunta de la asistencia a la emergencia de Ruanda realizada en 1996 (Borton et al., 1996) es una de las evaluaciones de acción humanitaria más influyentes jamás realizadas (Buchanan-Smith, 2003). El cómo y el por qué se estudiaron algunos años más tarde, concluyendo que el éxito de la evaluación podía explicarse en términos del modelo de los tres círculos ([Figura 3.2](#)):

- Fue una evaluación de alta calidad, debido a la minuciosidad y el rigor de la investigación, el alto calibre del equipo y la forma en que se formularon las recomendaciones. Tuvo una alta credibilidad entre los potenciales usuarios.





- En términos de cultura y estructura institucionales, el sistema humanitario internacional estaba abierto al cambio en ese momento, en parte debido al extendido malestar frente al desempeño altamente variable de las organizaciones humanitarias.
- Una serie de personas muy bien ubicadas hicieron campaña a favor del uso de la evaluación y crearon o explotaron hábilmente alianzas o redes para promocionarla.
- En términos de relaciones, los evaluadores tenían relaciones sólidas con los responsables de las políticas del sector.
- Las influencias externas también fueron importantes. Tras el escándalo de la crisis de Ruanda, y la intensa cobertura de prensa, los responsables de las políticas sectoriales estuvieron más abiertos al cambio.

**Ejemplo de buena práctica: Evaluación de 2009 de la asistencia humanitaria alemana por el Ministerio Federal de Relaciones Exteriores y el Ministerio Federal de Cooperación Económica y Desarrollo**

En 2009, el Ministerio Federal de Relaciones Exteriores y el Ministerio Federal de Cooperación Económica y Desarrollo emprendieron una evaluación independiente de la asistencia humanitaria alemana con el fin de extraer conclusiones que pudieran usarse para fines de gestión. El proceso atrajo mucha atención, por tratarse de la primera vez en que los dos ministerios acometían una evaluación conjunta y también porque por primera vez se sometió la asistencia humanitaria oficial de Alemania a una evaluación externa.

Concluida en 2011, la evaluación ha tenido importantes efectos en la arquitectura de la asistencia humanitaria alemana. A ello contribuyeron diversos factores. En primer lugar, la evaluación coincidió con una reforma radical del sistema humanitario oficial. Los nuevos líderes designados se comprometieron con los resultados de la evaluación y los utilizaron en este proceso. Los hallazgos informaron la redacción de la primera estrategia humanitaria del Ministerio de Relaciones Exteriores, que lanzó una iniciativa de calidad sin precedentes, la misma que dio lugar a numerosas actividades relacionadas con los temas abordados por la evaluación.

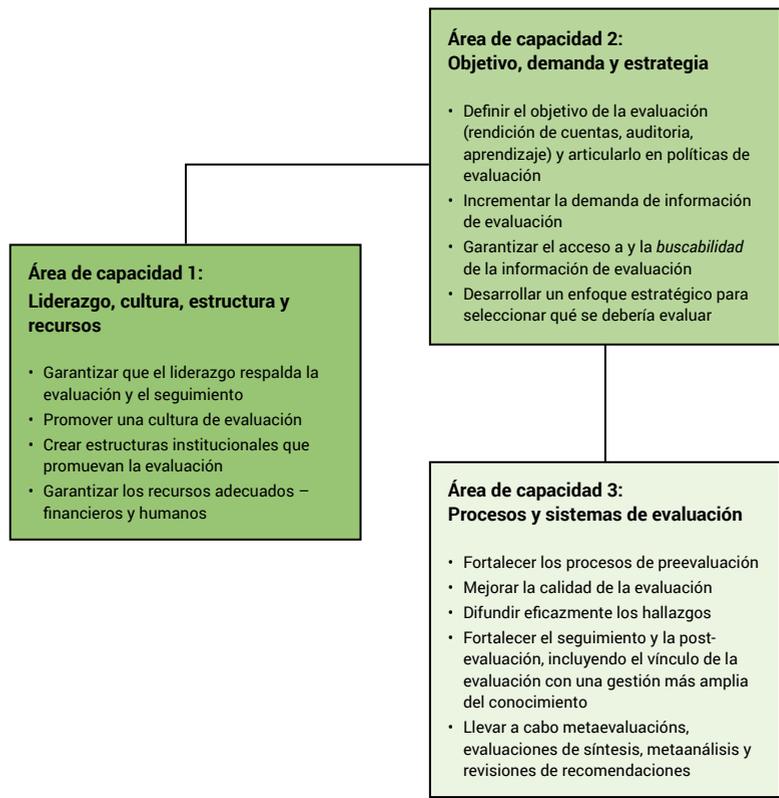
La recomendación de la evaluación de priorizar contrapartes, países y sectores más fuertes se ha mantenido vigente en lo que se refiere a la forma en que se decide la asistencia humanitaria oficial. Por ejemplo, Relaciones Exteriores también está invirtiendo en formas de identificar áreas de trabajo prioritarias en términos de sectores y países.

Fuente: Weingärtner et al. (2011)

## El marco de áreas de capacidad de evaluación de ALNAP

Hallam y Bonino (2013) proponen un marco analítico simple para facilitar el análisis y la discusión de los factores que influyen en si las evaluaciones generan un impacto dentro de la organización. Como explican los autores: El “pensamiento evaluativo” debería ser premiado y debería impregnar a toda la organización. No obstante, la gente debe estar dispuesta a cambiar, y recibir una recompensa por ello, por las mejoras que realizan en sus programas. Necesitan contar con las herramientas y el tiempo para analizar lo que hacen, así como explorar los resultados de una manera que fomente el cambio. La organización en su conjunto tiene que aceptar que esto supondrá un desafío para las formas existentes de trabajo y un debate importante. El liderazgo y la cultura son de gran importancia, como lo son los sistemas formales e informales y las estructuras de recompensa. (Hallam y Bonino, 2013: 18-19)

Figura 3.3 Marco de áreas de capacidad de evaluación de ALNAP



# 4 / Tipos de evaluación

Una vez que hayan decidido quiénes son los usuarios primarios previstos, lo que estos quieren y deben saber y el propósito general de la evaluación en términos de rendición de cuentas y/o aprendizaje, acto seguido deberán decidir el tipo de evaluación (FICR, 2010: 15-16).

## 4.1 ¿Qué tipo de evaluación van a emprender?

Para identificar el tipo de evaluación que van a acometer, deberán plantearse las siguientes preguntas:

1. **¿Cuáles son los alcances de su evaluación?**  
¿Está centrada en el nivel de proyecto, programático, sectorial, temático o global?
2. **¿En qué nivel de resultados están interesados?**  
¿Desean evaluar el proceso, los efectos o el impacto? ¿Hasta dónde desean llegar en la cadena de resultados? Para más detalles véase la Sección 5: Enmarcando su evaluación.
3. **¿En qué momento de la ejecución de la intervención se realizará la evaluación?**  
¿La evaluación está pensada para influir en la programación en una etapa temprana, por ejemplo, se trataría de una evaluación en tiempo real (ETR)? ¿O estamos pensando en una evaluación de término medio? ¿O la evaluación tendrá lugar una vez concluido el programa (expost)? El momento dependerá en gran medida del propósito general de la evaluación; es decir, si está orientada al aprendizaje o a la rendición de cuentas.
4. **¿Cuántos actores participan en la evaluación?**  
¿Se trata de una evaluación realizada por una sola agencia, de una evaluación conjunta con participación de uno o más actores o de una evaluación global del sistema humanitario para someterlo a revisión?
5. **¿Quiénes participan en la ejecución de la evaluación?**  
La mayoría de las EAH son realizadas por equipos que no han participado en la ejecución del programa o de intervención, o bien por equipos mixtos, compuestos por evaluadores externos e internos. En ocasiones, los involucrados en la ejecución del programa realizan una evaluación propia o autoevaluación. Las evaluaciones también pueden llevarse a cabo de manera participativa e involucrar a la población afectada de principio a fin, pero este caso es extremadamente raro en las EAH.

## 6. ¿Hay algún otro aspecto distintivo de la evaluación?

Por ejemplo, ¿la evaluación se centra en una política específica o se trata de una metaevaluación?

Estas categorías no son mutuamente excluyentes, y las EAH tradicionalmente pertenecen a varias categorías a la vez. Por ejemplo, la evaluación de un programa (alcance) podría ser una evaluación de impacto (nivel de resultados) que se lleva a cabo de manera ex-post (momento temporal) como una evaluación conjunta (con participación de varios actores) y de manera independiente (qué partes interesadas la llevan a cabo). El tipo de evaluación o la combinación de varios tipos contribuyen también a determinar el diseño de evaluación, así como el tipo de preguntas de evaluación apropiadas.

Las respuestas a las preguntas antes señaladas deben reflejarse claramente en los términos de referencia (TdR) de la evaluación.

## 4.2 Definición de diferentes tipos de evaluación

### Evaluaciones con distintos alcances



#### **Definición: Evaluación de proyectos**

Evaluación de una única intervención humanitaria, con objetivos específicos, recursos y un cronograma de ejecución específicos, que a menudo existe en el contexto de un programa más amplio

Un ejemplo de este tipo de evaluación es la evaluación de los proyectos de alojamiento de Tearfund y Tear Holanda tras el terremoto de 2009 en Padang, Sumatra (Goyder, 2010). En ocasiones se puede evaluar un grupo pequeño de proyectos relacionados entre sí y ejecutados por una misma organización, como en el caso de los proyectos financiados por ECHO en Pakistán y ejecutados por NRC (Murtaza et al., 2013).



#### **Definición: Evaluación de programas**

Evaluación de un conjunto de intervenciones que comparten un objetivo humanitario unificador

Un ejemplo de este tipo es la evaluación de la asistencia de Danida a las personas en situación de desplazamiento interno en Angola (Cosgrave, 2004). En ocasiones se reúne un grupo de intervenciones con objetivos comunes en un solo programa, como en este ejemplo, sin que las intervenciones hayan sido diseñadas para constituir un programa desde un inicio, como en el caso del programa de educación de NRC en Palestina (Shah, 2014).

**Definición: Evaluación de clusters**

Evaluación de múltiples proyectos pertenecientes a un programa más grande, o evaluación referida al Sistema de Coordinación de Clusters de Naciones Unidas

Dos ejemplos de una evaluación de clusters son la evaluación de AusAid de un clúster de ONG después del terremoto de Pakistán (Crawford et al., 2006) y la evaluación del trabajo de las ONG en la región del Pacífico (Crawford e Eagles, 2008).

El término evaluación de clusters se refiere cada vez más a las evaluaciones del Sistema de Coordinación de Clusters de Naciones Unidas, entre ellas la evaluación inicial de Stoddard et al. (2007), la Evaluación del Clúster Global de Logística o la Evaluación del Clúster Global de Seguridad Alimentaria (Steets et al., 2013) y del Clúster de Nutrición en la respuesta a las inundaciones de 2010 en Pakistán (Clúster de Nutrición, 2011).

**Definición: Evaluación de contrapartes**

Evaluación de un conjunto de intervenciones ejecutadas por una sola organización

La evaluación de Norad del trabajo de NRC es un ejemplo de este tipo de evaluación (Ternström et al., 2013), al igual que las evaluaciones de ECHO de sus contrapartes, tal como su asociación con el Organismo de Obras Públicas y Socorro de las Naciones Unidas (Grünewald y De Geoffroy, 2009).

**Definición: Evaluación sectorial**

Evaluación de un grupo de intervenciones que pertenecen a un mismo sector, las cuales contribuyen al logro de un objetivo humanitario específico. La evaluación puede abarcar una parte de un país, un único país o varios (UNICEF, 2013).

Un ejemplo de este tipo de evaluación es la evaluación del sector de agua, saneamiento e higiene en la RDC (Van der Wijk et al., 2010). Para algunas organizaciones, las evaluaciones sectoriales han sido recientemente desplazadas por el segundo tipo de evaluación de clusters que vimos arriba. Un ejemplo sería la evaluación de la FICR del clúster de alojamiento y artículos no alimentarios en la respuesta al ciclón Giri de 2010-2011 en Myanmar (Hedlund, 2011).



#### **Definición: Evaluación temática**

Evaluación de una selección de intervenciones que inciden en una prioridad humanitaria específica en varios países, regiones y posiblemente agencias y sectores

Un ejemplo de una evaluación temática es la evaluación del papel de la ayuda alimentaria en personas refugiadas en situación de desplazamiento prolongado (Cantelli et al., 2012) o la evaluación del diagnóstico de necesidades tras el tsunami del Océano Índico en diciembre de 2004 (De Ville de Goyet y Morinière, 2006). Se trata de uno de los cinco estudios temáticos realizados para la Coalición de Evaluación del Tsunami. La revisión realizada por el Alto Comisionado de las Naciones Unidas para los Refugiados (ACNUR) sobre el sector educación es un ejemplo de evaluación temática de un solo sector (Dryden-Peterson, 2011).



#### **Definición: Evaluación de un conjunto de proyectos humanitarios**

Evaluación del conjunto de proyectos humanitarios de una organización

Las evaluaciones de este tipo son efectuadas por entidades donantes para revisar el conjunto de sus proyectos humanitarios. Algunos ejemplos son la evaluación de la acción humanitaria finlandesa (Telford, 2005), la acción humanitaria suiza (De Ville de Goyet et al., 2011) y la estrategia humanitaria danesa (Mowjee et al., 2015).

## Evaluaciones por niveles de resultados



#### **Definición: Evaluación de impacto**

Evaluación enfocada en los efectos más generales de un programa humanitario, entre ellos el impacto intencionado y fortuito, el impacto positivo y negativo y el impacto a nivel macro (sectorial) y micro (de hogares, individual)

Un buen ejemplo de una evaluación de impacto es la evaluación de la asistencia otorgada por la Oficina de Población, Refugiados y Migración del Departamento de Estado de Estados Unidos a los refugiados de Burundi (Telyukov et al., 2009). Otro excelente ejemplo es la evaluación de impacto de los esfuerzos de reconstrucción impulsados por la comunidad en el condado de Lofa en Liberia (Fearon et al., 2008). Las evaluaciones de impacto son tratadas con mayor detalle en la [Sección 18](#).

**Definición: Evaluación de proceso**

Evaluación centrada en los procesos por los que los insumos se transforman en productos. Asimismo, puede examinar la intervención en su conjunto.

Las evaluaciones de proceso no buscan evaluar los efectos o el impacto sino los procesos que subyacen a la acción humanitaria. Por ejemplo, ACNUR evaluó su relación con, y uso de, el Fondo Central de Respuesta a Emergencias (CERF), con la finalidad de establecer cómo este había afectado el financiamiento de los programas de ACNUR y contribuido a las tendencias de financiamiento en un período de cinco años (Featherstone, 2014).

**Definición: Evaluación normativa**

Evaluación que compara las acciones que se están ejecutando con las planificadas o bien con estándares específicos

Las evaluaciones normativas son relativamente inusuales en la EAH. Estas evaluaciones se hacen para evaluar el cumplimiento de un conjunto de estándares normativos y no deben confundirse con las evaluaciones del trabajo normativo, donde se evalúa el trabajo de establecer normas y estándares (UNEG, 2013). Un ejemplo de una evaluación normativa fue la evaluación del Comité de Emergencia para Desastres (DEC) de la respuesta a la crisis de 2002-2003 en África Austral (Cosgrave et al., 2004), en la cual se examinó si la respuesta se atuvo al *Código de Conducta Relativo al Socorro en Casos de Desastre para el Movimiento Internacional de la Cruz Roja y la Media Luna Roja y las Organizaciones No Gubernamentales* (SCHR y FICR, 1994).

## Evaluaciones en momentos diferentes

**Definición: Evaluación en tiempo real**

Evaluación de una operación humanitaria en ejecución mientras se desarrolla

Entre los ejemplos, cabe destacar la evaluación de la FICR de la respuesta al tifón Haiyan en Filipinas (Greenhaigh et al., 2014) y la evaluación de ACNUR de su respuesta a la crisis en Siria (Crisp et al., 2013).



**Definición: Evaluación de término medio**

Evaluación que se lleva a cabo a mediados del periodo de implementación de una intervención

Entre los ejemplos se incluye la revisión de término medio del proyecto de diagnóstico de necesidades de Siria (Featherstone, 2013) y la evaluación de término medio de la Dirección General de Asistencia Humanitaria y Protección Civil de la Comisión Europea (ECHO) de la Decisión Regional sobre la Sequía en el Cuerno de África (Wilding et al., 2009).



**Definición: Evaluación ex-post**

Evaluación que se lleva a cabo una vez que la intervención ha concluido

Un ejemplo es la evaluación de CARE de su respuesta de emergencia al tsunami en dos distritos de Sri Lanka (Bhattacharjee et al., 2007).



**Definición: Evaluación continua**

Un conjunto de evaluaciones diseñadas para llevarse a cabo en el transcurso del tiempo de vida de una intervención

Algunos ejemplos de evaluaciones continuas son las dos evaluaciones del IASC de la respuesta al terremoto de Haití, realizadas entre tres y 20 meses después del sismo, respectivamente (Grünewald et al., 2010; Hidalgo y Théodate, 2012), así como una evaluación de seis años de duración de ACNUR al término de una intervención (Skran, 2012).



**Definición: Evaluación ex ante**

Evaluación que se lleva a cabo antes de iniciar una intervención

Estas evaluaciones, poco habituales en el sector humanitario, se basan en las lecciones aprendidas de operaciones similares. Un ejemplo es la revisión de la propuesta de crear un Cuerpo Voluntario Europeo de Ayuda Humanitaria (Bruaene et al., 2010).

## Evaluaciones según el número de actores



### **Definición: Evaluación de una sola agencia**

Evaluación del trabajo realizado por la agencia que ejecutó la intervención

Ejemplos de este tipo de evaluación, muy común en la EAH, son la evaluación de CARE de su respuesta al terremoto de Pakistán (Kirkby et al., 2006) y la evaluación de Oxfam Gran Bretaña del modelo de transferencias en efectivo en Sudán (Bush y Ati, 2007).



### **Definición: Evaluación conjunta**

Evaluación que realizan dos o más agencias para evaluar su trabajo conjunto

Las evaluaciones conjuntas pueden implicar un donante y un organismo receptor, múltiples agencias con misiones similares o diferentes actores que trabajan en el mismo sector. Algunos ejemplos son las evaluaciones conjuntas realizadas por un consorcio de ONG de un proyecto de desarrollo de capacidades de emergencia tras el terremoto de Yogyakarta (Wilson et al., 2007) y la evaluación conjunta de CARE y Save the Children de su respuesta al terremoto de Haití (O'Hagen et al., 2011). Entre las evaluaciones conjuntas de donantes se encuentra la evaluación conjunta expost del proyecto para vincular la ayuda de emergencia, la rehabilitación y el desarrollo (Brusset et al., 2013). Ejemplos del sistema de Naciones Unidas incluyen la evaluación conjunta del Programa Mundial de Alimentos y ACNUR sobre el impacto de la ayuda humanitaria en las personas refugiadas en desplazamientos prolongados (Cantelli et al., 2012).



### **Definición: Evaluación global del sistema humanitario**

Evaluación de la respuesta del sistema humanitario internacional a una crisis humanitaria, abierta a todos los actores del sistema

Ejemplos de este tipo de evaluación son la evaluación de la Coalición de Evaluación del Tsunami (TEC) de la respuesta internacional al tsunami del Océano Índico de 2004 (Telford et al., 2006) y la evaluación conjunta de la respuesta a la crisis de Ruanda (Borton et al., 1996).

## Evaluaciones llevadas a cabo por diferentes actores

**Definición: Evaluación externa o independiente**

Evaluación efectuada por evaluadores que no forman parte del equipo ejecutor

Las evaluaciones externas o independientes suelen ser un requisito de las evaluaciones orientadas a la rendición de cuentas por aportar un punto de vista objetivo.

**Definición: Evaluación propia o autoevaluación**

Evaluación realizada por las personas responsables de diseñar y ejecutar una intervención; también llamada evaluación interna

La mayoría de las evaluaciones propias se enfocan en el aprendizaje de lecciones aprendidas más que en la rendición de cuentas y suelen generar documentos de uso interno. Un ejemplo publicado es la evaluación de la respuesta de Medair al tsunami del Océano Índico de 2004 en Sri Lanka, realizada por una persona miembro del equipo inicial de Sri Lanka que regresó posteriormente para llevar a cabo la evaluación (Lee, 2005).

**Definición: Evaluación participativa**

Evaluación en la cual las partes interesadas, entre ellas la población afectada, trabajan juntas en el diseño, ejecución e interpretación de los resultados de una evaluación (ver diseño participativo en Sección 11)

Las evaluaciones participativas son muy poco comunes en el sector humanitario. Según King (2005), existe la percepción equivocada de que basta con involucrar al personal, los beneficiarios o los participantes de un programa de alguna manera – la más común consiste en recabar información de ellos – para que una evaluación sea “participativa”. El simple contacto entre un evaluador y los receptores o participantes de un programa no es suficiente, advierte. Lo que hace “participativa” a una evaluación es el rol que cumplen las partes interesadas, entre ellas los participantes del programa, así como la extensión de su participación a lo largo del proceso de evaluación (Alexander y Bonino, 2014).

## Otros tipos de evaluación



### **Definición: Evaluación de la tecnología**

Evaluación de una técnica o tecnología en especial

Estas evaluaciones no son comunes. Algunos ejemplos incluyen la revisión del uso del Módulo de Saneamiento Masivo de la Cruz Roja Británica tras el terremoto de Haití en 2010 (Fortune y Rasal, 2010) y la revisión del kit de materiales de albergue distribuido tras las inundaciones de Nigeria en 2012 (Bravo et al., 2014). Estas evaluaciones pueden centrarse asimismo en la evaluación de enfoques innovadores como el uso del modelo de gestión comunitaria de la desnutrición aguda en Nepal (Guerrero, 2010).



### **Definición: Evaluación institucional**

Evaluación de la dinámica interna de las organizaciones ejecutoras y de sus instrumentos de política, sus mecanismos de prestación de servicios y sus prácticas de gestión, así como de los vínculos entre ellos

Algunos ejemplos incluyen las dos evaluaciones del programa de fortalecimiento de capacidades humanitarias de UNICEF, financiado por el Departamento para el Desarrollo Internacional del Reino Unido (DFID, por sus siglas en inglés) (Bhattacharjee et al., 2010; Brown et al., 2005), y la revisión independiente de la respuesta operativa de UNICEF al terremoto de Haití en enero de 2010 (Bhattacharjee et al., 2011).



### **Definición: Evaluación de políticas**

Evaluación que examina las percepciones, creencias y supuestos que hacen posible, y deseable, la puesta en marcha de proyectos particulares. Pueden evaluar tanto la eficacia de la política en sí como la forma en que dicha política ha sido puesta en práctica.

Las evaluaciones de políticas a menudo son muy amplias y tienden a ser exámenes en lugar de evaluaciones. Un ejemplo es la revisión de ACNUR de su política con personas refugiadas en zonas urbanas en Nueva Delhi (Morand y Crisp, 2013). En ocasiones, dichas revisiones de políticas abarcan a todo el sector humanitario, como la Revisión de la Respuesta Humanitaria de las Naciones Unidas (Adinolfi et al., 2005) y la Revisión de la Respuesta Humanitaria a Situaciones de Emergencia del DFID (Ashdown, 2011).



### **Definición: Metaevaluación**

Evaluación diseñada para consolidar los hallazgos de múltiples evaluaciones, o evaluación de la calidad de una o más evaluaciones

Un ejemplo es la metaevaluación de evaluaciones conjuntas de ALNAP (Beck y Buchanan-Smith, 2008). Scriven (2011) ha diseñado una lista de verificación para metaevaluación del segundo tipo mencionado en la definición. Otro ejemplo es la metaevaluación de la calidad de las evaluaciones de Estados Unidos sobre la ayuda internacional extranjera (Kumar y Eriksson, 2012), al igual que los informes del Sistema Mundial de Vigilancia de los Informes de Evaluación de UNICEF (Universalía, 2013). UNICEF ha publicado los estándares con base en los cuales se juzgan dichas evaluaciones (Oficina de Evaluación de UNICEF, 2013). Para un análisis más detallado del concepto de metaevaluación, ver la [Sección 17.6](#).

## 4.3 Evaluaciones en tiempo real (ETR)



### **En profundidad: Evaluaciones en tiempo real (ETR)**

**Objetivo principal:** Proporcionar retroalimentación de manera participativa, durante el trabajo de campo, a los responsables de ejecutar y coordinar una respuesta humanitaria (Cosgrave et al., 2009b)

**Uso:** En los últimos años, el número de ETR en el sector de la acción humanitaria ha experimentado un incremento excepcional, en tanto que las agencias tratan de potenciar la contribución de estas evaluaciones al aprendizaje, en especial al aprendizaje inmediato que se produce durante la ejecución.

#### **Ventajas de las ETR:**

1. **Respecto a los plazos** – Las ETR se realizan en coyunturas de toma de decisiones clave, tanto operativas como de política. Pueden sacar a la luz importantes aspectos que han sido pasados por alto en la prisa por atender necesidades inmediatas. Por ejemplo, la ETR de Catholic Relief Services (CRS) en Pakistán en 2010 se llevó a cabo apenas nueve semanas después del inicio de la respuesta de la organización e incluyó una jornada de reflexión con el personal y las contrapartes en diferentes ubicaciones, en el transcurso de las cuales se elaboraron planes de acción inmediatos (Hagens e Ishida, 2010).



2. **Facilitan la interacción tanto entre los miembros del personal como con los evaluadores** – Las ETR suponen un diálogo continuo con el personal, tanto a nivel de los programas de campo como de la oficina central, y pueden brindar un canal de comunicación entre ambos grupos de personal, obviando obstáculos burocráticos. Por ejemplo, la ETR de 2010 de la Respuesta del Comité Permanente Interagencial (IASC) a las inundaciones de Pakistán incluyó una segunda visita, en el transcurso de la cual se discutieron los hallazgos y las conclusiones y recomendaciones con las partes interesadas como parte de un ejercicio participativo (Polastro et al., 2011b).
3. **Perspectiva** – El equipo de evaluación de una ETR puede abordar una emergencia desde diversos ángulos, al hablar con miembros del personal de todos los niveles, en diferentes países, con la población afectada y con las agencias socias y funcionarios del gobierno. Esto les permite formarse una opinión menos mediatizada por la resolución de los problemas cotidianos.

**El calibre de los evaluadores:** Los miembros del equipo de evaluación tienen que contar con la experiencia operativa necesaria para comprender los productos, efectos e impactos más probables de las acciones en curso.

**Desafíos y posibles soluciones:**

- Terminar el informe de la ETR antes de abandonar el terreno, a diferencia de una evaluación convencional.  
*Incorporar horas de trabajo para la redacción del informe en el cronograma del trabajo de campo.*
- Las partes interesadas primarias de una ETR son el equipo de campo y los responsables de coordinar la evaluación desde la oficina central, la mayoría de los cuales se encuentra bajo intensa presión y tienen poco tiempo para leer un informe de evaluación convencional.  
*Incluir en el plan de evaluación reuniones sobre el terreno para informar sobre el proceso, por ejemplo, una reunión hacia la mitad del trabajo de campo, para hablar con el equipo en terreno sobre los problemas que se hayan ido presentando, así como una reunión al final.*

## 4.4 Evaluaciones conjuntas



### En profundidad: Ventajas de la evaluación conjunta<sup>1</sup>

**Objetivo principal:** La evaluación realizada por varias agencias se lleva a cabo con la finalidad de determinar su impacto colectivo y/o promover el aprendizaje colectivo y/o evaluar determinados aspectos de la relación interinstitucional. El propósito declarado de una EAH conjunta suele ser tanto la rendición de cuentas como el aprendizaje; pero en la práctica, las organizaciones participantes, por ejemplo las afiliadas a la Coalición de Evaluación del Tsunami, han destacado, a menudo, que son más adecuadas para fines de aprendizaje.

**Uso:** Las evaluaciones conjuntas están cobrando cada vez mayor popularidad en el sector humanitario, lo cual refleja una mayor tendencia hacia la "acción conjunta".<sup>1</sup> Podemos encontrar evaluaciones conjuntas de las fuentes de financiamiento multidonantes, por ejemplo la evaluación del apoyo a los esfuerzos de paz de Sudán del Sur (Bennett et al., 2010); en consorcios de ONG, destaca, por ejemplo, la evaluación interinstitucional de la respuesta al terremoto de Yogyakarta (Wilson et al., 2007) y la evaluación de Acción Conjunta de las Iglesias (Alianza ACT) de la respuesta al terremoto de Haití (McGearty et al., 2012); así como en el concepto de la Agenda de Transformación de Naciones Unidas, por ejemplo las ETR del IASC (Cosgrave et al., 2010; Grünewald et al., 2010; Polastro et al., 2011a). Podemos encontrar evaluaciones conjuntas asimismo en evaluaciones globales del sistema como la evaluación de la respuesta al tsunami del Océano Índico en 2004 (Telford et al., 2006).

#### Ventajas:<sup>2</sup>

1. **Enfocadas en un panorama más amplio** – Las evaluaciones conjuntas invitan a enfocarse en el panorama general y en cuestiones estratégicas y con un enfoque más orientado a políticas.
2. **Aptas para evaluar el impacto** – Las evaluaciones conjuntas evitan atribuir el impacto a una sola agencia y pueden utilizarse para evaluar el impacto colectivo de una respuesta humanitaria.
3. **Mayor trayectoria de consulta con la población afectada** – Al disponer de mayores recursos que la mayoría de las evaluaciones de una sola agencia, las evaluaciones conjuntas tienen una mayor trayectoria de consulta a la población afectada (ver Beck y Buchanan-Smith, 2008).
4. **Utilización** – Las evaluaciones conjuntas a menudo gozan de mayor credibilidad y pueden ser aprovechadas para hacer incidencia política a favor de los cambios deseados. Además, pueden cumplir una valiosa función de aprendizaje interinstitucional, ayudando a las organizaciones participantes a aquilatar sus respectivos enfoques e intercambiar buenas prácticas.



5. **Financieras y de gestión** – Las evaluaciones conjuntas comparten recursos y capacidades de evaluación y pueden reducir los costos de transacción al disminuir el número de evaluaciones de una sola agencia consultando a las mismas partes interesadas, en las zonas afectadas.

**Calibre de los evaluadores:** Las evaluaciones conjuntas complejas requieren una combinación de capacidades técnicas, políticas e interpersonales, especialmente del jefe del equipo.

La cantidad de evaluadores que reúnen dichas competencias es reducida. Planificar y convocar a concurso desde un inicio ayudará a disponer de una reserva de candidatos calificados.

**Desafíos y posibles soluciones:**

- Debido al número de partes interesadas que participan en una evaluación conjunta, la negociación de los TdR tiende a tomar mucho más tiempo que en una evaluación de una sola agencia.  
*Es importante programar tiempo adicional para la negociación de los TdR. Las destrezas de dirección y moderación son fundamentales en esta fase inicial de planificación, para poder manejar las negociaciones y cerciorarse de que los TdR finales sean claros y concisos, y no una larga lista de aspectos sin ningún orden ni priorización.*
- La complejidad de las estructuras de gestión y los altos costos asociados a una evaluación conjunta pueden constituir una barrera para la participación de organizaciones más pequeñas, por ejemplo ONG nacionales.  
*Identificar opciones para involucrar a ONG más pequeñas, disminuyendo las exigencias – por ejemplo convocándolas solo en momentos clave de la evaluación y/o exigiéndoles aportes económicos menores que a las organizaciones más grandes – ayudará a asegurar su participación.*

# 5 / Enmarcando su evaluación

Esta sección se refiere a cómo darle un marco a su evaluación. “Enmarcar” significa establecer o identificar una estructura que ayude a traducir las preguntas de evaluación generales en preguntas específicas. Este proceso puede basarse tanto en supuestos sobre cómo se supone que la intervención generará el cambio deseado (preguntas causales) como en los estándares que rigen dichas intervenciones (preguntas normativas).

La selección de un marco causal o normativo dependerá del contexto. En ocasiones resulta más apropiado focalizar la atención en preguntas normativas, por ejemplo cuando resulta imposible establecer la atribución y es difícil evaluar la contribución.

En esencia, la idea detrás de enmarcar una evaluación es que los evaluadores puedan identificar los aspectos más significativos y concentrarse en ellos, antes que en los que resultan menos importantes para que la intervención logre su objetivo.

**¿Cómo garantizar la calidad de su evaluación?** Un enfoque de control de calidad cuidadosamente planificado y estructurado ayuda. Existen varios momentos fundamentales para someter a control de calidad una evaluación. Con demasiada frecuencia, el énfasis principal del control de calidad recae sobre el informe de evaluación. Una opción mucho más sensata consiste en enfocarse en el control de calidad en una etapa más temprana del proceso.

A lo largo de toda la guía identificaremos los momentos del proceso de evaluación que resultan más adecuados para que los gestores de la evaluación evalúen la calidad de los productos y adapten el proceso a partir de los resultados.

Estas oportunidades clave para el control de calidad se presentan a lo largo de una curva en “S”, ya que el nivel de influencia del gestor de una evaluación sobre la calidad de la misma disminuirá a medida que se desarrolla el proceso de evaluación.



## Oportunidades de control de calidad: Alcances y enfoque de evaluación



El primer momento clave para realizar el control de calidad es durante el proceso de formulación de su evaluación – que incluye determinar el alcance y el enfoque de su evaluación.

## 5.1 La lógica del programa

En la fase de diseño de una intervención, las acciones programadas idealmente deben basarse en una teoría que explique cómo se alcanzará el resultado final deseado. Las teorías explícitas suelen ser presentadas como un tipo de modelo lógico.



### Definición: Modelo lógico

Cuadro o diagrama donde se presenta la teoría del programa (la forma en que se espera que los insumos contribuyan al objetivo general) para una intervención

Los modelos lógicos van desde marcos lógicos hasta diagramas y teorías del cambio complejas (para un buen resumen general, ver Funnell y Rogers, 2011).

## Marcos conceptuales

Los marcos conceptuales son modelos lógicos, generalmente sustentados en una investigación exhaustiva, que ilustran los factores más importantes de un tema concreto. Esta base de investigación significa que los marcos conceptuales son mucho más robustos que otros modelos lógicos que se presentan en esta sección. Mientras que otros modelos lógicos pueden sustentarse en la presunción de nexos causales, los marcos conceptuales se basan en nexos establecidos como resultado de una investigación.

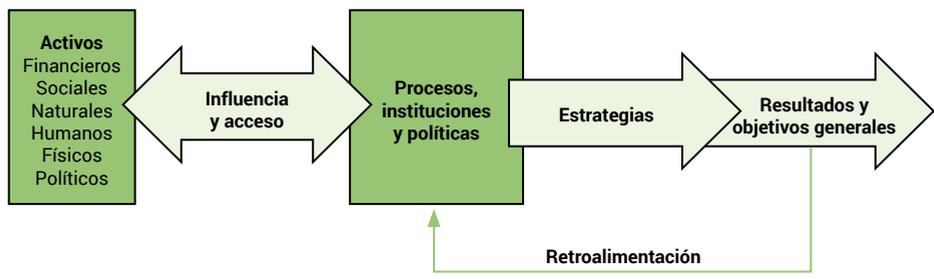


### Consejo útil

Verifique si el tema de su evaluación se sustenta en un marco conceptual existente. Los marcos conceptuales son herramientas sumamente útiles para cerciorarse de que ningún área importante sea pasada por alto.

El marco de medios de vida sostenibles examina los medios de vida en el medio rural. Lautze y Raven-Roberts (2003: 10) propusieron la variante que aparece a continuación para emergencias humanitarias complejas.

Figura 5.1: Herramienta de análisis de medios de vida



Fuente: Lautze y Raven-Roberts (2003: 10)

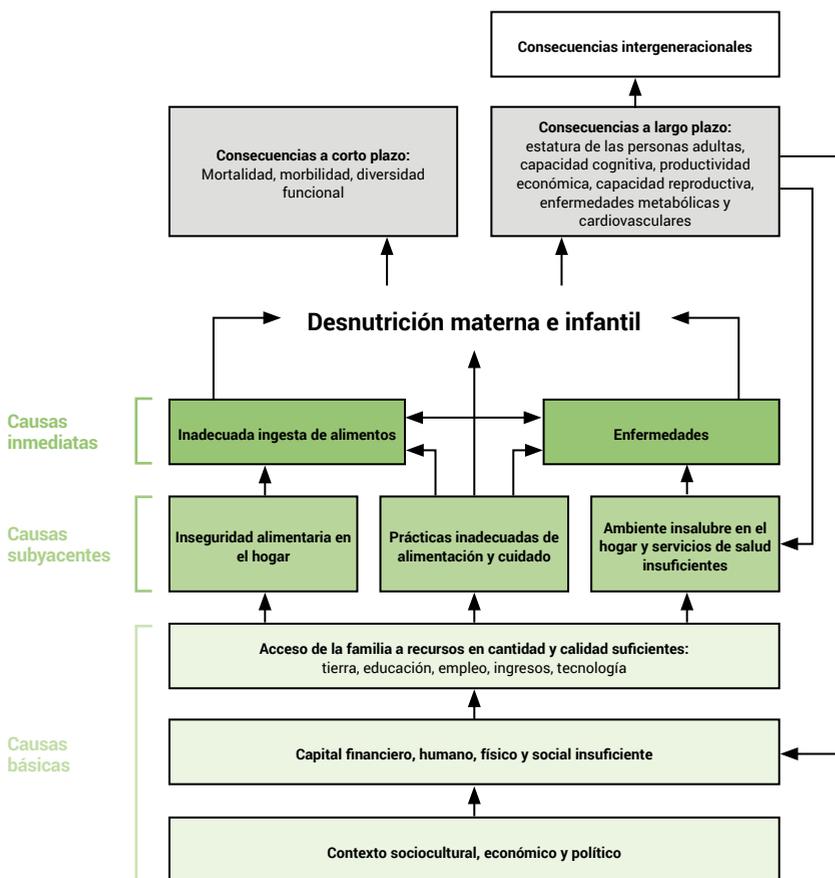
El equipo de evaluación utilizó este marco para la evaluación del programa de la FAO en Somalia (Buchanan-Smith et al., 2013). Esto supuso que, si bien el equipo examinó la forma en que la FAO apoyaba los medios de vida mediante la entrega de activos, en especial la distribución de medicinas y alimento para ganado a las asociaciones de ganaderos, también analizó la actuación de la organización en la esfera de las políticas públicas. Un problema en este ámbito era el veto al ganado de la región establecido por los países de Medio Oriente, que tuvo enormes repercusiones en los medios de vida de la población.

Otro modelo conceptual que puede ser particularmente útil para evaluar la acción humanitaria es el marco conceptual de UNICEF sobre las causas de la desnutrición, especialmente si el foco de la evaluación recae sobre los programas de ayuda alimentaria o intervenciones médicas.

La evaluación de Oxfam GB del programa de Seguridad Alimentaria y Medios de Vida Urbanos en situaciones de emergencia (Macauslan y Phelps, 2012) utilizó este modelo conceptual como base para el marco causal de la desnutrición y sus efectos en los medios de subsistencia.

Incluso si no se hubiera utilizado un marco conceptual en el desarrollo de una intervención, los evaluadores aún pueden hacer uso de un marco conceptual ya que este contiene nexos causales probados. El marco conceptual también puede sugerir preguntas de evaluación importantes. Por ejemplo, al evaluar un programa de nutrición no sorprendería que los resultados fueran negativos si no se hubiesen tomado en cuenta aspectos como la salud o el cuidado de la infancia.

Figura 5.2: Marco conceptual de UNICEF sobre las causas de la desnutrición infantil



Fuente: UNICEF (2013)

## Marcos lógicos

Los marcos lógicos son una de las formas más sencillas de modelo lógico, que vinculan las actividades directamente con los resultados. El concepto fue introducido por primera vez en el sector de desarrollo en 1970 por USAID (Rosenberg et al., 1970), para ser adoptado más ampliamente después. El Cuadro 5.1 presenta los elementos más importantes del marco lógico. Las agencias pueden usar nombres diferentes y definiciones concretas para los elementos de una cadena de resultados.

Cuadro 5.1: Muestra de marco lógico

Jerarquía de resultados	Indicadores	Supuestos
<b>Meta u objetivo general</b> El objetivo de más alto nivel al que se espera que contribuya el proyecto	Indicadores para medir los avances hacia el logro del objetivo de desarrollo	Supuestos relacionados con la sostenibilidad del objetivo de desarrollo
<b>Propósito u objetivo específico</b> El efecto que se espera generar como resultado del proyecto	Indicadores para medir los avances hacia el logro del propósito u objetivo específico	Supuestos relacionados con el logro del objetivo de desarrollo dado que el propósito u objetivo específico ha sido alcanzado
<b>Productos</b> Los resultados que la realización del proyecto debería poder garantizar	Indicadores para medir el grado de generación de los productos	Supuestos relacionados con el logro del propósito u objetivo específico una vez que están elaborados los productos
<b>Actividades</b> Las acciones emprendidas para generar los productos	Insumos, tales como los bienes y servicios requeridos para realizar las actividades	Supuestos relacionados con la elaboración de los productos

Fuente: Adaptado de Norad (1999)

## Cadena de resultados

El enfoque del marco lógico ha sido criticado por su excesiva rigidez, especialmente en contextos humanitarios rápidamente cambiantes. Además, la jerarquía de los resultados puede prestarse a confusión ya que el objetivo de desarrollo de un proyecto puede ser el mismo que el propósito del programa del cual forma parte.

Norad ha ido adoptando un enfoque más flexible para el uso de una cadena de resultados, tal como se demuestra a continuación.

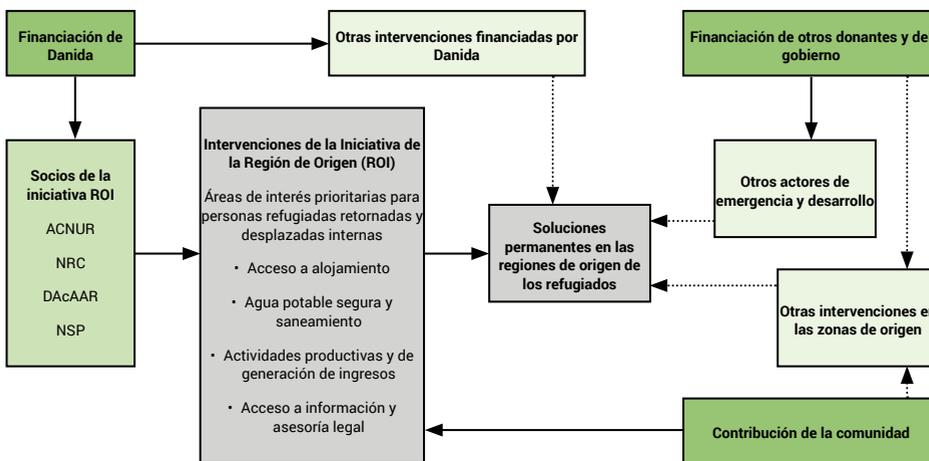


La cadena de resultados puede tener nombres diferentes, entre ellos el de rutas causales, pero la idea principal es mostrar cómo se pretende que los insumos generen el cambio a largo plazo deseado.

## Modelos lógicos gráficos

Los modelos también pueden ser presentados en forma de gráficos, como en el siguiente ejemplo de una evaluación de la iniciativa danesa de las regiones de origen en Afganistán. Una presentación gráfica como esta puede ayudar a sacar a la luz los nexos con actividades de terceros.

Figura 5.3: Evaluación de la asistencia danesa a Afganistán



La Fundación Kellogg ha publicado una guía para el desarrollo de modelos lógicos (Fundación W.K. Kellogg, 2004), que se enmarca en cadenas de resultados bastante simples.



### Consejo útil

Los modelos son simplificaciones adecuadas de una realidad compleja. Esta simplificación constituye su mayor fortaleza, ya que permite la manipulación de conceptos y situaciones complejas, pero también su mayor debilidad, porque puede inducirnos a pasar por alto algún factor fundamental. Si utiliza un modelo en una validación, examine hasta qué punto el modelo se ajusta a la situación real que está examinando y modifíquelo de acuerdo a ello de ser necesario.

Los paquetes de *software* como *DoView* pueden facilitar el trazado de modelos lógicos simples como *pipelines* o cadenas de efectos.

## 5.2 Teoría del cambio



### Definición: Teoría del cambio

Descripción del mecanismo central mediante el cual se produce un cambio en las personas, grupos y comunidades

Una teoría del cambio (TdC) expresa las conexiones o cadenas de conexiones causa-efecto entre una intervención y el resultado final deseado. Este enfoque ha adquirido cada vez más popularidad, y algunos donantes han adoptado la TdC en vez de modelos lógicos (aunque algunos requieren ambos). Vogel señala que las TdC son vistas como “una herramienta de análisis más realista y flexible que los enfoques actuales basados en el marco lógico” (2012: 9). La guía de Keystone sobre la TdC compara la TdC y el marco lógico (2009, 30), mientras que Rogers (2012: 7) incluye los marcos lógicos en un listado de cuatro opciones para representar una TdC.

Las teorías del cambio son especialmente útiles en entornos complejos, ya que permiten desagregar la cadena de resultados en una serie de conexiones causales que pueden ser validadas por la evaluación.

Sin embargo, el consenso en torno a la TdC es aún menor que respecto a otras formas de modelo lógico. Stein y Valters (2012: 5) consideran problemático que diferentes organizaciones utilicen el término TdC para referirse a cosas muy distintas.

Una TdC puede ser representada en forma de enunciado (ver ejemplos en el CAD de la OCDE, 2012: 85), de cuadro o de diagrama. En el ejemplo que ofrecemos a continuación, de la evaluación interinstitucional de la asistencia alimentaria a los refugiados de larga data en Bangladesh, la TdC se presenta en forma de tabla como un “modelo lógico simplificado” (Nielsen, 2012: 4).

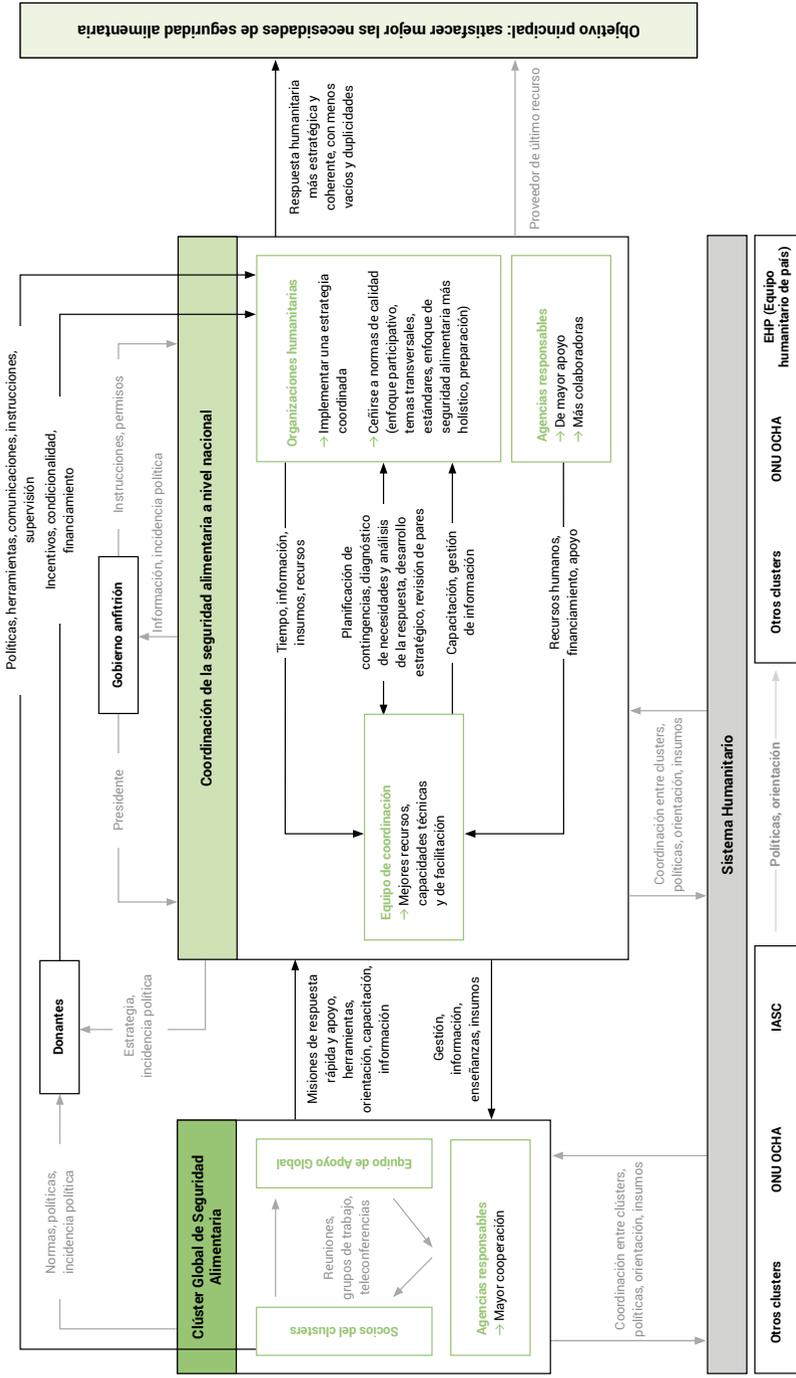
**Cuadro 5.2:** Ejemplo de evaluación interinstitucional de la asistencia alimentaria para los refugiados de larga data

Tiempo	Asistencia alimentaria	Supuestos	Resultados esperados
<b>Corto plazo</b>	Distribución general de alimentos – raciones completas	Asistencia en la modalidad de respuesta a una situación de emergencia	Vidas salvadas; consumo de alimentos mejorado; seguridad y protección provistas. Nivel mínimo de autosuficiencia.
<b>Mediano plazo</b>	Disminución de la asistencia alimentaria (raciones parciales)	Transición después de la respuesta a la emergencia; disponibilidad de intervenciones complementarias a nivel de servicios sociales, por ej. agua, saneamiento, educación, vivienda, etc.	Canasta de alimentos mejorada, nivel nutricional mejorado (desnutrición aguda y crónica).  Mayor capacidad de la población afectada para generar medios de vida.
<b>Largo plazo</b>	Disminución de la asistencia alimentaria (raciones parciales)	Disponibilidad de intervenciones de apoyo a medios de vida; construcción de activos	Autosuficiencia de los refugiados; integración local; reubicación o repatriación

El informe inicial de la evaluación del Clúster de Seguridad Alimentaria elaboró un gráfico de la TdC (Steets et al., 2013: 6), tal como se muestra en la [Figura 5.4](#).

Entre los recursos disponibles sobre el desarrollo de una TdC se encuentran Funnell y Rogers (2011) y CARE Internacional (2012). Asimismo, la revisión de teorías de cambio de Comic Relief de 2011 ofrece un buen resumen. El anexo de la guía del CAD de la OCDE sobre la evaluación de intervenciones de construcción de la paz incluye una evaluación para entender y evaluar una TdC (OCDE/CAD, 2012: 80-86)

Figura 5.4: Clúster de Seguridad Alimentaria – Teoría del cambio



## 5.3 Desarrollo de un modelo lógico retroactivo

Idealmente, la intervención debería haber sido diseñada alrededor de una TdC para generar un documento claro sobre cómo debería funcionar. Sin embargo, esto todavía no constituye una práctica común, y el resultado final es que los TdR pueden incluir una cláusula exigiendo a los evaluadores que desarrollen una TdC explícita para que se adecue a la TdC implícita en la intervención.

En principio, una TdC implícita estuvo en la base de las decisiones en el caso del modelo lógico de la Iniciativa de la Región de Origen en Afganistán (ROI) de Danida arriba descrita, pero la omisión de explicitar la TdC puede acarrear problemas. Por ejemplo:

- Es poco probable que una evaluación que abarca más de un proyecto o programa se sustente en una única lógica general. Ese fue el caso en la evaluación de la estrategia de evaluación humanitaria danesa (Mowjee et al., 2015: 18).
- Uno de los aspectos más útiles de una TdC explícita es que provee un marco para que los evaluadores validen su puesta en práctica. Si los evaluadores desarrollan la TdC, el riesgo es que utilicen la misma información para generar la teoría tanto como para validarla.
- Si los evaluadores encuentran que una cadena causal en particular (tal como fue postulado en la TdC que ellos mismos desarrollaron) es inválida, la razón podría ser que es su modelo, y no la intervención, el que es inválido.

Una TdC desarrollada retroactivamente apenas puede servir de débil remplazo para una TdC desarrollada antes de que se inicie un programa y que sea utilizada para guiarlo. Desarrollar una TdC después de una intervención se asemeja un poco a tratar de hacer planificación estratégica después de los hechos. Si tienen que desarrollar una teoría del cambio ex post, una manera de acometer esta tarea es convocando a un taller de partes interesadas para elaborar la TdC implícita bajo la cual operó la intervención, o bien validar una TdC explícita que los evaluadores pudieran haber elaborado con base en el estudio documental.

## 5.4 Marcos de evaluación y criterios de evaluación

El segundo tipo de enmarcación se realiza en función de los estándares normativos que rigen la acción humanitaria. Los contextos humanitarios suelen ser complejos y existe el riesgo de pasar por alto algún aspecto fundamental, que la evaluación adolezca de falta de rigor conceptual o que no cumpla con los estándares aceptados. Los marcos normativos pueden ayudar con estos problemas al:

- Brindar una estructura para desagregar la tarea de la evaluación en elementos más pequeños y por lo tanto más fáciles de manejar.
- Reducir el riesgo de que elementos fundamentales sean pasados por alto, al dirigirse la atención del equipo sistemáticamente hacia todos los elementos del sujeto de evaluación.
- Brindar una estructura para la evaluación que sea reconocida por las partes interesadas. Las partes interesadas pueden haber tenido poco contacto con criterios de evaluación convencionales, pero a menudo están familiarizadas con la estructura de los marcos conceptuales o con las normas de uso común en su sector.
- Servir de línea de base de buenas prácticas universalmente aceptadas a partir de las cuales validar el proyecto.

Los criterios del CAD de la OCDE que se describen en la [Sección 6, Selección de las preguntas de evaluación](#), representan un posible marco para una evaluación, si bien todos los marcos son más adecuados para verificar las preguntas que para formularlas.

Otros marcos pueden añadir especificidad y detalle a la evaluación y hacerla más accesible a los usuarios. Dichos marcos pueden utilizarse adicionalmente a los criterios del CAD de la OCDE o incluso sustituirlos, si son apropiados. Con frecuencia se pide a los evaluadores que utilicen uno de los siguientes tipos de marcos:

- Marcos normativos amplios que establecen las normas que definen el esfuerzo humanitario. Entre estos se encuentran el derecho internacional humanitario y los principios que guían la acción humanitaria (principios humanitarios), así como diversas convenciones internacionales.
- Estándares y guías, que pueden usarse no solo como normas a partir de las cuales evaluar las intervenciones sino también como un medio para desagregar las intervenciones humanitarias en componentes que sean más fáciles de examinar. Entre estos se encuentran los estándares sistémicos y sectoriales y las guías y manuales institucionales.

En el [Cuadro 5.3](#) encontrarán un resumen de los marcos de evaluación y sus respectivos usos. Este listado no es exhaustivo, y constantemente se están desarrollando nuevos marcos. El marco seleccionado debe calzar con el contexto.

Cuadro 5.3: Ejemplos de marcos de evaluación y sus respectivos usos

Marcos de evaluación y ejemplos	Posibles usos
<p><b>Marcos normativos amplios</b></p> <ul style="list-style-type: none"> <li>• El derecho internacional humanitario, por ejemplo el Protocolo Relativo a la Protección de las Víctimas de Conflictos Armados sin Carácter Internacional (CICR, 1977)</li> <li>• Los Principios Humanitarios (Wortel, 2009)</li> <li>• La Convención sobre los Derechos del Niño (Asamblea General de las Naciones Unidas, 1989)</li> <li>• La Convención sobre el Estatuto de los Refugiados (Asamblea General de las Naciones Unidas, 1951)</li> </ul>	<p>Los marcos normativos pueden emplearse en la fase inicial para determinar si hay algunos aspectos de la intervención que pueden provocar inquietud y necesitan ser examinados más detenidamente. Algunos marcos también proveen listas de verificación o estándares a partir de los cuales revisar las políticas y el desempeño.</p>
<p><b>Normas y directrices para todo el sistema</b></p> <ul style="list-style-type: none"> <li>• Principios y buenas prácticas en la donación humanitaria (Good Humanitarian Donorship, 2003)</li> <li>• Código de Conducta para las ONG, la Cruz Roja y la Media Luna Roja (Borton, 1994)</li> <li>• Norma humanitaria esencial en materia de calidad y rendición de cuentas (CHS Alliance, 2015)</li> <li>• Principios rectores de los desplazamientos internos (OCHA, 2004)</li> <li>• Principios para el compromiso internacional en estados frágiles (OCDE/CAD, 2007)</li> <li>• Norma Humanitaria Esencial en materia de calidad y rendición de cuentas (HAP, 2014)</li> <li>• COMPAS Calidad (Groupe URD, 2009)</li> </ul>	<p>Estas normas pueden servir como lista de verificación o punto de referencia para evaluar el desempeño, como base para desagregar la evaluación en tareas individuales más manejables y como estructura para el informe. Son particularmente eficaces cuando una organización es signataria formal de los acuerdos; de otro modo, justificar su uso puede ser difícil. La fase inicial es el momento más propicio para proponer la aplicación de una norma en particular.</p>
<p><b>Normas y estándares sectoriales</b></p> <ul style="list-style-type: none"> <li>• Elementos sectoriales específicos de las Normas Mínimas del Proyecto Esfera (Proyecto Esfera, 2011)</li> <li>• Normas Mínimas para la Educación en Situaciones de Emergencia (INEE, 2006)</li> <li>• Normas y Directrices para Intervenciones Ganaderas en Emergencias (Proyecto LEGS, 2009)</li> <li>• Normas Mínimas para la Recuperación Económica (Red SEEP, 2010)</li> <li>• Normas Mínimas para la Protección de la Infancia en la Acción Humanitaria (Grupo de Trabajo para la Protección de la Infancia, 2012)</li> </ul>	<p>Las normas sectoriales constituyen una buena base para organizar evaluaciones sectoriales. En algunos casos se sustentan en normas generales. Por ejemplo, la norma mínima del Proyecto Esfera sobre la consulta con las poblaciones afectadas está diseñada para aplicar a todos los sectores.</p>



Marcos de evaluación y ejemplos	Posibles usos
<p><b>Normas o estándares y guías institucionales</b></p> <ul style="list-style-type: none"> <li>Manual para Situaciones de Emergencia del Alto Comisionado de las Naciones Unidas para los Refugiados (ACNUR, 2007)</li> <li>Manual para la Evaluación de la Seguridad Alimentaria en Emergencias del Programa Mundial de Alimentos (PMA, 2002)</li> <li>Manual para Situaciones de Emergencia sobre el Terreno de UNICEF (UNICEF, 2005)</li> </ul>	<p>Estos documentos pueden proveer una buena base para verificar el cumplimiento (la rendición de cuentas) y también para desagregar y organizar la evaluación.</p>
<p><b>Normas y guías nacionales</b></p> <ul style="list-style-type: none"> <li>Leyes, normas y políticas nacionales</li> <li>Lineamientos acordados a nivel nacional (por ejemplo, a nivel de clusters)</li> </ul>	<p>Estos documentos a menudo se basan en modelos internacionales más amplios, pero son adaptados al contexto local. Por lo tanto, suelen ser más apropiados para el contexto que las normas y guías internacionales más generales.</p>



#### Consejo útil

Sigan la pista de los TdR y los documentos del proyecto. Si los TdR y los documentos de planificación del proyecto aluden a marcos o estándares conceptuales específicos, podrían utilizar uno de ellos como marco para la evaluación.

En ocasiones surgen problemas cuando un informe de evaluación intenta utilizar varios marcos a la vez. Esto puede llevar a un informe largo, ya que cada sección del marco A tendrá subsecciones para el marco B (y a veces hasta para el marco C).



#### Consejo útil

Utilicen un solo marco para estructurar el informe. Si una evaluación utiliza un marco y este es central para la evaluación, podría proveer una estructura valiosa para el informe. Si la evaluación utiliza diversos estándares, sería conveniente escoger uno como marco principal y tratar los otros en anexos cortos del informe principal.

# 6 / Selección de las preguntas de evaluación

La presente sección se centra en la selección de las preguntas de evaluación. Estas deben basarse en lo que sus usuarios primarios previstos necesitan saber para marcar una diferencia en su trabajo, tal como se explica en la [Sección 3, Pensar desde el principio y a menudo en el uso de una evaluación](#). Las preguntas pueden ser divididas de la siguiente manera:

- Preguntas de más alto nivel, por ejemplo “¿Cuán eficaz fue nuestra respuesta?”
- Preguntas de evaluación en sí, que se derivan de las preguntas de más alto nivel
- Preguntas que son planteadas a las personas entrevistadas, grupos focales y sujetos de encuestas en entrevistas, guías temáticas e instrumentos de encuestas

Esta última categoría se aborda en la [Sección 13, Metodologías de campo](#).

## 6.1 Impacto de la selección de las preguntas en la calidad de la evaluación

La selección de las preguntas de evaluación tiene un efecto decisivo en la calidad de la evaluación.

Los siguientes factores pueden llevar a una evaluación deficiente:

- Demasiadas preguntas, lo cual quiere decir que la evaluación no podrá profundizar en ningún tema;
- Preguntas que no se ajustan a los alcances y el enfoque de la evaluación;
- Preguntas que serían mejor respondidas por otros medios (por ejemplo, una auditoría o un estudio temático);
- Preguntas que no están encontrando respuestas utilizables.

## Puesto de control de calidad: Preguntas de evaluación



Las preguntas que formulan determinarán qué diseños y métodos son factibles y cuáles son los más apropiados. Las preguntas determinan asimismo qué presupuesto se requiere. De manera fundamental, la selección de las preguntas tiene una fuerte influencia en el uso de una evaluación.

## No todas las evaluaciones hacen preguntas

La evaluación participativa del programa de escuelas de campo para agricultores de Lira, Uganda, realizada en 2008, no incluyó ninguna pregunta, limitándose a los objetivos de la evaluación. El evaluador elaboró luego un cuadro con los requisitos de información necesarios para cumplir con los objetivos (Foley, 2009). En la evaluación de la intervención danesa en Somalia no se incluyeron preguntas, y se solicitó a los evaluadores “profundizar y especificar las preguntas de evaluación para los temas presentados arriba” en una matriz de evaluación (Gartner et al., 2012: 72). Si bien este enfoque sigue siendo relativamente inusual, ha sido adoptado actualmente por la Unión Europea como una práctica regular en sus evaluaciones.

## Las preguntas de evaluación son el motor de toda la evaluación

Las preguntas de evaluación cumplen un papel decisivo en el sentido de que definen la evaluación. Vale la pena hacer el esfuerzo de seleccionar las preguntas adecuadas para garantizar la calidad de la evaluación, ya que las preguntas:

- Determinan el tipo de evaluación que se elija, por ejemplo, evaluación de impacto, evaluación de proceso, ETR, etc.
- Determinan el diseño o diseños más apropiados. Por ejemplo, si la pregunta se refiere al efecto de la asistencia, posiblemente deban realizar un estudio comparativo, como las evaluaciones de impacto conjuntas de la ayuda alimentaria y el estudio de caso sobre las crisis de refugiados de larga duración en Ruanda (Sutter et al., 2012).
- Determinan las metodologías que deben ser utilizadas. Por ejemplo, si en los TdR dice que la evaluación debe “establecer el nivel de satisfacción general de los beneficiarios” es posible que deban realizar una encuesta, como en el caso de la evaluación de la FICR sobre el programa de alojamiento transitorio en Indonesia (Tango, 2009: 38).
- Determinan la generación de opciones sobre la matriz de evaluación. Esto se analiza en mayor detalle en la Sección 8, Fase inicial. Ver Buchanan et al. (2009: 37-39) en el que se incluye un ejemplo de cómo las preguntas definen la matriz de evaluación.
- Deberían determinar el presupuesto de la evaluación. Por ejemplo, las preguntas que solo pueden ser respondidas mediante una encuesta elevarán el costo de la evaluación.
- Influyen en el tipo y la magnitud de las decisiones de muestreo. La evaluación de UNICEF de su programa de educación en situaciones de emergencia y de transición posteriores a la crisis empleó una estrategia de muestreo intencional para cumplir con los objetivos del estudio (Barakat et al. 2010: 194).
- Determinan las recomendaciones. Como establece la guía del CAD de la OCDE para la evaluación de la ayuda al desarrollo, los hallazgos y las conclusiones de la evaluación son las respuestas a las preguntas planteadas y seleccionadas para la evaluación. Las lecciones aprendidas y las recomendaciones proveen el vínculo entre los resultados de la evaluación y el desarrollo futuro de las políticas y programas (1991:10).

## 6.2 Tipos de preguntas de evaluación

El Cuadro 6.1 presenta cinco tipos de preguntas de evaluación, tomados de Better Evaluation (2013) y de Morra Imas y Rist (2009: 223-229).

Cuadro 6.1: Tipos de preguntas de evaluación

Tipo de pregunta	Ejemplo	Comentario
Descriptiva	¿Cómo utilizó la población afectada los kits con materiales de alojamiento que le fueron entregados?	Este es un tipo de pregunta bastante común en la EAH. ¿Pero qué hacemos con los resultados?
Normativa	¿Hasta qué punto el alojamiento provisto cumplió con las normas mínimas del Proyecto Esfera?	Otras preguntas normativas pueden preguntar por el logro de las metas del proyecto o por el cumplimiento de las normas institucionales, o bien por la comparación con otras organizaciones. Están enfocadas en la <u>rendición de cuentas más que en el aprendizaje</u> .
Causal	¿Hasta qué punto la entrega de asistencia a las aldeas desalentó la migración a la capital regional?	Este es a menudo el tipo de pregunta más difícil de responder en una EAH, ya que es difícil evaluar la atribución o incluso la contribución en vista de la naturaleza caótica, compleja y complicada de una intervención humanitaria. Las preguntas sobre el <u>impacto son causales por definición</u> . Ver discusión de la causalidad en la <u>Sección 16, Análisis</u> .
Evaluativa	¿Fue apropiada nuestra política de distribuir kits con materiales de alojamiento solo a quienes pudieran probar que eran propietarios de un terreno?	Pide a los evaluadores que emitan un juicio de valor con respecto al programa. Ver la discusión sobre el razonamiento evaluativo en la <u>Sección 16, Análisis</u> . "Por definición, las evaluaciones responden preguntas evaluativas, es decir, preguntas relativas a la calidad y el valor" (Davidson, 2014: 1). No obstante, la mayoría de las EAH van más allá y exigen recomendaciones: los juicios de valor no generan recomendaciones en sí mismos.
Orientada a la acción	¿Cómo podríamos ayudar mejor a las personas vulnerables a reconstruir sus albergues?	Las preguntas orientadas a la acción son particularmente apropiadas para las EAH, ya que vinculan la pregunta directamente con las acciones que la organización podría tomar en el futuro.

Las preguntas orientadas a la acción son complejas, y el equipo de evaluación normalmente debe contestar varias subpreguntas para responderlas. En el ejemplo de arriba sobre opciones para ayudar mejor a las personas vulnerables a reconstruir sus albergues, las subpreguntas implícitas serían:

- ¿Cuáles son las características del proyecto? (descriptiva)
- ¿Quiénes clasifican como vulnerables? (normativa)
- ¿Qué quiere decir ayudar mejor? (evaluativa)
- ¿Qué factores ayudan a conseguir un mejor apoyo? (causal)

Solamente después de responder a estas subpreguntas, el equipo de evaluación podrá hacer un listado de opciones o formular recomendaciones.



#### **Consejo útil**

Siempre que puedan, enmarquen las preguntas de evaluación centrales como preguntas orientadas a la acción.

Las preguntas orientadas a la acción son muy apropiadas para la EAH porque, al igual que las subpreguntas correspondientes, están fuertemente enfocadas en cómo se utilizarán realmente los hallazgos. Además, las preguntas orientadas a la acción son una excelente elección para enmarcar las preguntas centrales.

El proceso mediante el cual los evaluadores responden las preguntas descriptivas, normativas, causales y evaluativas y luego formulan recomendaciones es a menudo un poco errático. Hacer preguntas orientadas a la acción significa que las recomendaciones no se dejan al azar o a merced de las preferencias personales de los evaluadores, sino que señalan claramente dónde quieren ustedes que les den orientación de cara a los programas futuros.



#### **Consejo útil**

Consideren pedirles a los evaluadores que les den varias opciones, en vez de una sola recomendación.

La toma de decisiones es competencia de la dirección. Pedir opciones quiere decir que los evaluadores debe valorar las ventajas y desventajas de cada recomendación, en vez de formular una sola. Esto no significa que los evaluadores no puedan identificar su opción u opciones preferidas, pero para la dirección puede ser más útil que se le presenten opciones. Se trata de un enfoque de cierta manera similar a pedir a los evaluadores que presenten las conclusiones y luego desarrollen recomendaciones en un taller de partes interesadas con los usuarios de la evaluación.

## 6.3 Número de preguntas

Los TdR de una evaluación a menudo plantean demasiadas preguntas. Si bien hay algunos ejemplos en que esto no sucede, como el caso de la evaluación de los efectos de la ayuda alimentaria para los refugiados de larga duración (Sutter et al., 2012), que solo planteó cuatro preguntas de evaluación, no es inusual que los TdR identifiquen 20 preguntas o más.

En una revisión de la calidad de las evaluaciones del DFID, Manning recomendó limitar el número de preguntas de evaluación “con el fin de evitar informes muy largos y discursivos” (2009: xi). En una revisión relacionada, Perrin halló que incluso las evaluaciones más superficiales del DFID debían responder decenas de preguntas, y que la “falta de focalización con frecuencia llevó a una atención superficial a demasiadas consideraciones, y al final con frecuencia a una larga lista de hallazgos donde los hallazgos e implicaciones centrales son difíciles de reconocer” (2009: vii).

El problema de hacer muchas preguntas de evaluación es aún peor para las EAH, que a menudo se ejecutan con un presupuesto limitado.

Existen varias razones para limitar el número de preguntas:

1. **Centrar la evaluación.** Como apuntó Perrin (2009: vii) en su revisión de las evaluaciones del DFID, “Cuando todo es prioritario... nada lo es”. USAID (1996: 2) observa asimismo: “Hacer demasiadas preguntas puede resultar en un esfuerzo descentrado”. El número de preguntas de evaluación debería por lo tanto circunscribirse a las áreas de interés prioritarias. Si hay muchas preguntas, el equipo de evaluación puede priorizar aquellas que le interesan más, que pueden no coincidir con las prioridades de la organización que encarga la evaluación.
2. **Asegurar que el equipo de evaluación ahonde lo suficiente en las preguntas.** Si hay 25 preguntas, los evaluadores solo podrán dedicar un tiempo limitado al abordaje de cada una. Esto lleva a informes de evaluación poco profundos.
3. **Aumentar la utilidad de la evaluación.** Pocas organizaciones pueden lidiar con más de media docena de recomendaciones a un nivel en particular. Cada pregunta de evaluación necesita una respuesta, y las respuestas suelen ser presentadas como hallazgos, los mismos que llevan a conclusiones. Si bien una recomendación puede sustentarse en varias conclusiones, es más común que una conclusión lleve a varias recomendaciones. Demasiadas recomendaciones harán más difícil utilizar la evaluación.

Los departamentos de evaluación están empezando gradualmente a limitar el número de preguntas de evaluación con el objeto de evitar estos problemas (Riddle, 2009: 16).

Hallam y Bonino (2013) comentan que:

El CICR solía pedir a todas las partes interesadas que identificaran el rango de preguntas y temas que les gustaría ver incluidos en una evaluación.

El departamento de evaluación reformuló este pedido posteriormente, añadiendo el adjetivo “evaluables” para calificar las preguntas. Sin embargo, los alcances de las evaluaciones siguieron creciendo, hasta que el proceso se tornó inmanejable. Para paliar este efecto, actualmente el departamento de evaluación trata de enfocarse en solo tres preguntas clave por evaluación (pág. 62).

El *software* que utiliza NRC para generar los TdR de las evaluaciones tiene cabida para tres preguntas de evaluación solamente. Method Labs elaboró un listado de preguntas guía con miras a restringir los alcances de la evaluación a la recolección de información (Buffardi, 2015).

## 6.4 Desentrañando las preguntas centrales

¿Cómo debemos proceder para evaluar preguntas de alto nivel como “¿Estamos aplicando la estrategia adecuada?” Estas preguntas son difíciles de responder con rigurosidad.

Un enfoque para desentrañar dichas preguntas es considerar cómo podría ser utilizada la respuesta a la pregunta de alto nivel. Sin duda, la razón para hacer una pregunta de esta naturaleza es decidir si se mantiene o no la estrategia actual o se mejora de alguna manera. También puede haber un interés en comparar las estrategias con las de otras organizaciones.

Las preguntas subyacentes probablemente serían:

- ¿Cómo podemos mejorar nuestra estrategia actual? (orientada a la acción)
- ¿Cuáles son las ventajas o desventajas de nuestra estrategia actual en comparación con otras empleadas en contextos similares? (normativa)

En resumen, pregúntense lo que la organización que encarga la evaluación podría hacer con la respuesta a una pregunta y luego consideren si esta puede perfeccionarse para proveer una mejor base para la acción. Incluso la primera de estas preguntas subyacentes probablemente sea excesivamente amplia. La estrategia actual consistirá de una combinación de elementos.

Siguiendo con la idea anterior, las preguntas pueden ser:

- ¿Qué podríamos agregar, eliminar o modificar para aumentar el impacto o reducir los costos? (orientada a la acción)
- ¿Cómo se compara nuestra estrategia con las de otras organizaciones y/o otros contextos? (normativa)

En este caso tendría sentido abocarse primero a comparar la estrategia con estrategias adoptadas en otros contextos, ya que esto se presta para un estudio documental. El estudio documental puede generar líneas de investigación para el examen de elementos de la estrategia actual.

Este proceso genera preguntas que son mucho más fáciles de responder con rigurosidad. Las preguntas de evaluación subyacentes aún tienen que ser respondidas, porque comparar estrategias o alternativas entraña juicios evaluativos, pero los juicios están ahora enmarcados en acciones prácticas. Ver la Sección 16: Análisis, para una discusión de cómo responder las preguntas de evaluación.

En ocasiones se deja en manos del equipo de evaluación el desglose de las preguntas de alto nivel en preguntas de evaluación que se esfuerzan por responder, así como la presentación de las mismas en el informe inicial, a menudo en una matriz de evaluación. Sin embargo, la organización que encarga la evaluación puede estar mejor posicionada para determinar el posible uso de las respuestas a las preguntas de evaluación.

## 6.5 Criterios de evaluación



### Consejo útil

Las preguntas de evaluación vienen primero.

Los criterios de evaluación y otros marcos son herramientas útiles una vez que hayan decidido sus preguntas de evaluación. *Los criterios y marcos son buenos sirvientes pero malos patrones.* Eviten hacer un listado de sus criterios primero y escoger las preguntas correspondientes a cada criterio después, y recuerden mantenerse fuertemente enfocados en cómo se usarán los hallazgos y las recomendaciones.

Los seres humanos hacemos evaluaciones todos los días. Ya sea al comprar manzanas en un supermercado o al elegir un emparedado en la cafetería, nos formamos un juicio sobre las opciones disponibles a partir de nuestros criterios personales de calidad y valor. Una EAH formal emplea un enfoque mucho más estructurado pero esencialmente se basa en los mismos criterios.

En 1991, el CAD de la OCDE estudió los problemas más comunes observados en los proyectos de desarrollo y propuso cuatro criterios de calidad – pertinencia, eficacia, sostenibilidad e impacto – así como eficiencia (OCDE/CAD, 1991). Unos años después, adaptó estos criterios para la EAH en emergencias complejas (OCDE/CAD, 1999), añadiendo los criterios de cobertura y coherencia, sugiriendo la adecuación como alternativa a la pertinencia y la conectividad como alternativa a la sostenibilidad y proponiendo dos criterios nuevos: coordinación y protección. Estos criterios reflejaban los problemas más importantes observados en la acción humanitaria en la década de 1990.

Aunque los criterios del CAD de la OCDE no son perfectos, existen buenas razones para basarse en ellos a la hora de verificar sus preguntas de evaluación:

- El uso de criterios estandarizados hace mucho más fácil la metaevaluación (la extracción de lecciones de una amplia gama de evaluaciones);
- Los criterios estandarizados tienden a identificar debilidades comunes en la acción humanitaria, con base en la experiencia y la investigación;
- Las evaluaciones que utilizan criterios estandarizados facilitan el trabajo de los gestores de la evaluación y otros evaluadores.



#### Consejo útil

No necesitan cubrir todos los criterios. Sus preguntas de evaluación pueden aludir solamente a uno o dos de los criterios solamente.

En la fase de planificación de la EAH, primero identifiquen lo que necesitan saber y luego sitúenlo con respecto a los criterios de evaluación – y no al revés. Los criterios son herramientas para ayudarnos a pensar y pueden sugerir otras preguntas pertinentes. Formulen preguntas de evaluación solo si están dispuestos a hacer algo a partir de las respuestas, y utilicen solo aquellos criterios que tengan que ver con las preguntas que desean ver contestadas. Son las preguntas lo que importa; no los criterios.

Una pregunta de evaluación puede tocar varios criterios. Por ejemplo, la pregunta “¿Hasta qué punto nuestra estrategia de focalización nos permitió ofrecer una asistencia oportuna de manera eficaz en función de los costos?” toca los criterios de: cobertura (focalización), eficacia (oportunidad) y eficiencia (eficacia en función de los costos). La pregunta puede subdividirse aún más, pero cualquier pregunta sobre la eficacia o la eficiencia de la focalización inevitablemente incluirá el criterio de la cobertura.

En el Cuadro 6.2 se identifica el criterio principal para cada una de las preguntas de muestra, con comentarios y una indicación sobre qué acciones podría tomar una organización con base en las respuestas. Adicionalmente se ofrece una pregunta alternativa enfocada en el uso, que puede ayudar a hacer menos discursivos los informes de evaluación.



### Consejo útil

Utilicen preguntas orientadas a la acción para evaluaciones más pequeñas. Las preguntas orientadas a la acción son particularmente apropiadas para evaluaciones pequeñas porque inducen a focalizarse directamente en lo que una organización podría hacer para mejorar su desempeño futuro.



## En profundidad: Preguntas de evaluación y criterios asociados

Cuadro 6.2: Preguntas de evaluación y criterios asociados

Preguntas de muestra	Criterio principal	Definición del criterio
<p>¿Hasta qué punto ha satisfecho nuestro programa las necesidades inmediatas?</p> <p>¿Cómo podemos cerciorarnos de que nuestros kits de cocina respondan mejor a las necesidades locales?</p>	Adecuación	Grado al que las actividades humanitarias se adecúan a las necesidades locales, fortaleciendo con ello la apropiación, la rendición de cuentas y la relación precio-calidad. (Sustituye el criterio de pertinencia que se utiliza en la evaluación de programas de desarrollo.)
<p>¿Hasta qué punto logró sus objetivos el programa, entre ellos la entrega oportuna de asistencia de emergencia?</p> <p>¿Qué cambios podemos hacer para reducir la tasa de deserción de los receptores de suplementos alimenticios?</p>	Eficacia	Grado al que una actividad logra su cometido, o si se puede esperar que esto suceda con base en los productos.
<p>¿Qué tan eficaz en función de los costos fue nuestro programa de alojamiento?</p> <p>¿Cómo podemos reducir los tiempos de espera en el centro de salud?</p>	Eficiencia	Productos – cualitativos y cuantitativos – logrados como resultado de los insumos.
<p>¿Cuál ha sido el impacto del programa de transferencias en efectivo, positivo y negativo?</p> <p>¿Qué medidas podríamos adoptar para reducir el daño ocasionado por la recolección de leña de los refugiados?</p>	Impacto	Los efectos más amplios del proyecto – sociales, económicos, técnicos y ambientales – en los individuos, grupos de género y edad, comunidades e instituciones. Los impactos pueden ser intencionales y fortuitos, positivos y negativos, a nivel macro (sectorial) y micro (de hogares). (No es exactamente lo mismo que el “impacto” en la cadena de resultados.)



← Cuadro 6.2: Preguntas de evaluación y criterios asociados

Preguntas de muestra	Criterio principal	Definición del criterio
<p>¿Cómo ha afectado la distribución gratuita de vacunas para ganado en el enfoque de recuperación de costos de veterinarios comunitarios?</p> <p>¿Qué podemos hacer para evitar que la distribución de alimentos disuada a los agricultores de sembrar sus propios cultivos?</p>	Conectividad	Grado en el que actividades de respuesta a una emergencia a corto plazo se llevan a cabo en un contexto que toma en cuenta problemas interrelacionados y a más largo plazo. Sustituye el criterio de sostenibilidad que se utiliza en la evaluación de programas de desarrollo.
<p>¿Hasta qué punto han beneficiado las transferencias de dinero a los hogares más vulnerables?</p> <p>¿Cómo podemos asegurar que los grupos y personas marginadas también tengan acceso a subsidios para alojamiento?</p>	Cobertura	Grado al que importantes sectores de la población sometidos a penurias extremas fueron alcanzados por la acción humanitaria.
<p>¿Qué tan coherentes son las políticas institucionales de protección y cuáles son las implicaciones?</p> <p>¿Cómo podríamos abogar para que otros donantes tengan en cuenta los derechos humanos en las decisiones relativas al financiamiento?</p>	Coherencia	Grado en el que las políticas de seguridad, desarrollo, comerciales y militares, así como las políticas humanitarias, son coherentes y toman en cuenta los aspectos humanitarios y de derechos humanos. (Más enfocado en las políticas de las entidades donantes, pero se puede usar también en organizaciones individuales respecto a la coherencia de sus políticas institucionales.)
<p>¿Qué tan eficiente ha sido la coordinación de la distribución de alimentos en la región y con qué consecuencias?</p> <p>¿Cómo podemos reducir las brechas en el suministro de agua por las diferentes organizaciones que utilizan camiones cisterna?</p>	Coordinación	Grado en que las intervenciones de diferentes actores armonizan entre sí, promueven las sinergias y evitan las brechas, duplicidades y conflictos por recursos (a menudo se convierte en el criterio de eficacia).

Fuente: Adaptado de Beck (2006), donde se proporciona mayor información sobre el uso de los criterios del CAD en la evaluación

Eviten empezar con los criterios y luego seleccionar las preguntas para cada uno, tanto debido a las limitaciones de recursos como porque una EAH debe enfocarse en las necesidades de los usuarios.



Las preguntas deben guiar la evaluación humanitaria. Los criterios de evaluación sirven solo a modo de lista de verificación, para asegurar que ningún área problemática de envergadura haya pasado desapercibida. Perrin (2009: xi) recomendó que las evaluaciones del DFID se enfocaran “en un número limitado de prioridades y temas de evaluación centrales, utilizando los criterios del CAD de manera inteligente antes que mecánica, para ayudar a identificar un pequeño número de preguntas de evaluación que puedan ser realísticamente respondidas, tomando en cuenta el contexto, la disponibilidad de información y los alcances y recursos facilitados para la evaluación”.

Para asegurar la integridad y la comparabilidad de las evaluaciones, es posible que algunas organizaciones deseen demostrar qué criterios de evaluación han sido cubiertos por preguntas específicas. En algunos casos, una pregunta puede rozar varios criterios, pero no siempre es posible cubrir todos los criterios.

### Distribuir los términos de referencia

Cuando los términos de referencia (TdR) son distribuidos en busca de comentarios, el número de preguntas de evaluación crece a medida que cada departamento y/o cada parte interesada añade sugerencias. Cuanto más ampliamente se circule el documento, tanto mayor será el número de preguntas. Hallam y Bonino (2013) puntualizan que “las evaluaciones de alta calidad, que satisfacen necesidades de información reales... *umentan a su vez la demanda de evaluaciones*”. Las preguntas que se añaden después de circular los TdR con frecuencia no han sido bien meditadas y tienden a desvirtuar en vez de elevar la demanda de evaluaciones.

Las estrategias para controlar el número de preguntas incluyen:

- Establecer en los TdR los objetivos generales de la evaluación solamente y dejar en manos del equipo el desarrollo de las preguntas de evaluación en el informe inicial;
- Pedir a los revisores que jerarquicen las preguntas sugeridas, y que prioricen y justifiquen las preguntas adicionales que pudieran proponer;
- Pedir a quienes sugieren preguntas adicionales que expliquen cómo utilizarían las respuestas y cómo ello justificaría cualquier costo adicional.



### Consejo útil

Distribuyan los TdR en formato PDF y no como un documento editable, para disuadir a los destinatarios de agregar preguntas.

## 6.6 Selección de las preguntas más idóneas

Si la evaluación tiene un gran número de preguntas, deberán jerarquizarlas. Una forma de hacerlo es asignando una puntuación a las preguntas en función de diferentes factores. Estos pueden incluir la medida en que las respuestas a las preguntas satisfarán las necesidades de los usuarios.

Por ejemplo, pregunten si las preguntas:

- Son fundamentales para el objetivo de la evaluación;
- Tienen respuestas que pueden ser puestas en práctica de inmediato;
- Pueden ser respondidas sin lugar a equivocación por la evaluación;
- Solo pueden ser respondidas por una evaluación;
- Son centrales para su mandato principal;
- Podrían mejorar la calidad del servicio que prestan;
- Podrían reducir los costos per cápita de prestar un servicio;
- Afectarían o serían de valor para casi todas las partes interesadas.

Los factores, o el peso que ustedes les asignen, variarán de acuerdo a los objetivos de la evaluación. Este cálculo puede ser presentado en forma de rúbrica (ver el Cuadro 6.3). Lo importante es decidir con anticipación qué factores son importantes y clasificar las potenciales preguntas de acuerdo a ello.

**Cuadro 6.3: Rúbrica para evaluar las preguntas de evaluación**

Pregunta sobre el factor Aplicar la pregunta sobre el factor a su potencial pregunta de evaluación	Puntuación				Pesos* ¿Cuán significativa es la pregunta sobre el factor para su evaluación?	Puntuación para cada potencial pregunta (puntuación x peso)				
	Puntuación = 1	Puntuación = 2	Puntuación = 3	Puntuación = 4		Pregunta potencial 1	Pregunta potencial 2	Pregunta potencial 3	Pregunta potencial 4	Pregunta potencial 5
A ¿Si tuviésemos la respuesta a esta pregunta, podríamos...?	estar mejor informados de nuestro trabajo?	usaría para abogar a favor de nuestra forma de hacer las cosas?	aplicaría en programas futuros?	aplicaría de inmediato en programas que están actualmente en ejecución?						
B ¿Puede la evaluación responder esta pregunta sin lugar a equivocación con los recursos disponibles?	Es poco probable	Posiblemente	Probablemente	Definitivamente						
C ¿Podría ser respondida esta pregunta por otros medios?	Si, pero tenemos el presupuesto para una evaluación.	Una evaluación es el medio más conveniente para responder a esta pregunta.	Una evaluación económica para responder a esta pregunta.	No. Solo una evaluación podría responder esta pregunta de manera efectiva.						
D ¿Qué tan central es esta pregunta para nuestro mandato principal?	Es periférica para el mandato principal.	Es relevante para el mandato principal.	Esperamos que sea el núcleo de nuestro mandato principal en el futuro.	Es central para el mandato principal.						
E ¿Cuál es el potencial impacto de la respuesta a esta pregunta en la calidad de los servicios que ofrecemos?	Escaso impacto en la calidad de nuestros servicios	Cierto impacto en la calidad de nuestros servicios	Gran impacto en la calidad de nuestros servicios	Podría mejorar sustancialmente la calidad de los servicios que ofrecemos.						
F ¿Cuál es el potencial impacto de la pregunta en el costo de los servicios que ofrecemos?	Escaso impacto en el costo de los servicios que ofrecemos	Podría generarnos algunos ahorros.	Podría generarnos un gran ahorro en el costo por cápita.	Podría reducir significativamente el costo de los servicios que ofrecemos.						
G ¿Qué proporción de nuestro grupo-objetivo global se vería afectado por la respuesta a esta pregunta?	Pocos	Pocos, pero particularmente vulnerables.	Muchos, incluyendo particularmente vulnerables	Casi todos						
<b>Puntuación total para pregunta potencial</b>										

Nota: Las preguntas (A-G) pueden ser más o menos significativas para los potenciales beneficiarios y pueden recibir puntuaciones variables. Las preguntas de evaluación se asignan pesos diferentes.

# 7 / Términos de referencia y presupuesto de una evaluación

Esta sección incide en los términos de referencia (TdR)<sup>3</sup> de la evaluación. Los TdR se plasman en un documento que especifica la tarea de evaluación con suficiente nivel de detalle como para que el potencial equipo de evaluación entienda lo que se espera de ellos.

La presente sección trata brevemente asimismo sobre el proceso de presupuestación, no solo para que el gestor de la evaluación entienda lo que es potencialmente viable sino también para que el equipo de evaluación prepare una oferta.

## 7.1 Términos de referencia



### **Definición: Términos de referencia**

Los TdR son “una descripción general de los requisitos y expectativas de una evaluación. Contienen una declaración explícita de los objetivos de la evaluación, los roles y responsabilidades de los evaluadores y el cliente de la evaluación, así como los recursos disponibles para la evaluación” (Roberts et al., 2011: 2).

Los TdR representan la base del contrato entre el equipo de evaluación y la organización que encarga la evaluación.

Los TdR son fundamentales para establecer la calidad de la evaluación, ya que las decisiones sobre los alcances y el enfoque pueden tener un gran impacto en la utilidad del resultado.

Los TdR definen los alcances en términos de: elementos del proyecto o programa a ser evaluados, período de tiempo e intervenciones específicas con mayores probabilidades de arrojar respuestas utilizables. Los TdR también pueden especificar el enfoque de evaluación que ha de ser aplicado y el presupuesto que debería ser utilizado para responder las preguntas. Establece si las respuestas pueden obtenerse mejor focalizando la evaluación en el personal, las organizaciones contrapartes o la población afectada.

## 7.1.1 Los términos de referencia y el informe inicial

Davidson (2012) subraya que la evaluación es un proyecto conjunto de la organización que encarga la evaluación y los evaluadores.

Una evaluación de alta calidad, que valga la pena, factible, no solo depende de la competencia técnica y las capacidades de consulta efectivas del evaluador. Las decisiones que tome y las acciones que adopte (o no) el cliente son decisivas para determinar el valor de una evaluación para una organización. Una evaluación de alta calidad es producto de una interacción fructífera entre un cliente bien informado y un equipo de evaluación receptivo y con las calificaciones adecuadas.

El gestor de la evaluación tiene dos opciones para involucrar al equipo de evaluación en la negociación de los alcances de la evaluación.

Una aproximación consiste en considerar los TdR como un borrador hasta que el equipo de evaluación haya sido contratado. Los temas pendientes se deciden como parte del proceso de formalización del contrato. Se trata de una buena práctica, pero los sistemas financieros de algunas organizaciones no permiten este tipo de negociación.<sup>4</sup>

Otro enfoque consiste en que el equipo de evaluación recurra al informe inicial para proponer cómo abordará los TdR. El gestor de la evaluación y el equipo de evaluación pueden valerse del informe inicial como base para negociar los alcances definitivos de la evaluación. Una misión inicial puede contribuir a la calidad del alcance de la evaluación final.

La combinación de TdR e informe inicial define *qué* es lo que ha de ser evaluado y *por qué*, además de *cómo*. Es por eso que la lista de verificación de UNEG incide tanto en los TdR como en el informe inicial (UNEG, 2010b). El gestor de la evaluación puede ya sea definir algunos aspectos con antelación o dejar en manos del equipo de evaluación el abordaje de los mismos en el informe inicial.

**Cuadro 7.1:** Elementos indispensables y deseables en los TdR y en el informe inicial de una evaluación

<b>Elementos típicamente incluidos en los TdR</b>	<ul style="list-style-type: none"> <li>• Contexto</li> <li>• Propósito (cómo se usará la evaluación)</li> <li>• Objetivos</li> <li>• Criterios</li> <li>• Alcance</li> <li>• Destinatarios</li> <li>• Roles y responsabilidades</li> <li>• Hitos</li> <li>• Entregables</li> <li>• Contenido del informe inicial</li> </ul>
<b>Elementos que pueden ser incluidos en los TdR o el informe inicial</b>	<ul style="list-style-type: none"> <li>• Marco de evaluación</li> <li>• Preguntas de evaluación</li> <li>• Fuentes que se utilizarán</li> <li>• Matriz de evaluación</li> <li>• Diseños de evaluación</li> <li>• Metodologías de recolección de información</li> <li>• Indicadores a ser evaluados</li> <li>• Metodologías de análisis de información</li> <li>• Contenido del informe de evaluación</li> </ul>
<b>Elementos típicamente incluidos en el informe inicial</b>	<ul style="list-style-type: none"> <li>• Plan de trabajo</li> <li>• Asignación de tareas a los miembros del equipo</li> <li>• Herramientas de recolección de información</li> </ul>

Los TdR transmiten al equipo de evaluación la intención del gestor de la evaluación, y el informe inicial es la respuesta del equipo a ello. Así, los TdR y el informe inicial configuran una especie de diálogo, complementado por conversaciones en torno al informe inicial. Los informes iniciales son apropiados incluso para evaluaciones pequeñas. La única excepción se da cuando la unidad o departamento encargado de elaborar los TdR es el que llevará a cabo la evaluación.

Como regla general, salvo en circunstancias excepcionales (por ejemplo, si el equipo de evaluación es relativamente inexperto), es mejor dejar que el equipo encare la mayor cantidad posible de temas en el informe inicial, en vez de definirlos uno mismo en los TdR. La razón de ello es que, para cuando se llega al informe inicial, los temas a ser evaluados habrán sido identificados con mucha mayor claridad. Asimismo, es mejor dejar que el equipo de evaluación presente el producto deseado.

Los TdR deben ser elaborados incluso para evaluaciones internas, de modo tal que el objetivo, el propósito y el alcance se encuentren claramente establecidos. Esto evita los malentendidos después.

## 7.1.2

### ¿Qué incluyen los TdR?

Incluso para la evaluación más pequeña, los TdR deben incluir lo siguiente (ver también el [Cuadro 7.1](#)):

- El contexto de la evaluación
- Los objetivos específicos de la evaluación
- Los alcances, en términos del proyecto o programas cubiertos, los sectores a ser evaluados, la extensión geográfica y el período cubierto
- Los hitos, las fechas límite y los entregables
- Los roles y responsabilidades del equipo de evaluación y del gestor de la evaluación
- Qué recursos, entre otras cosas qué información existente, se encuentran disponibles

En la mayoría de los casos, los TdR incluirán también las preguntas de evaluación que deben ser respondidas. Algunos TdR establecen únicamente los objetivos de la evaluación y dejan que sea el equipo el que proponga las [preguntas](#) (que deben ser aprobadas por el gestor de la evaluación) en el [informe inicial](#).

Para mayor orientación sobre la elaboración de los TdR ver los siguientes documentos:

- Plantilla de términos de referencia del DFID, un documento breve (2 páginas) y enfocado (DFID, 2010);
- Lista de verificación de la calidad de UNEG para TdR e informes iniciales (UNEG, 2010);
- Guía para la elaboración de términos de referencia del Banco Mundial (Roberts et al., 2011), que ofrece orientación exhaustiva;
- Guía para la redacción del documento “Alcance de trabajo” de USAID (USAID, 1996).

### Las evaluaciones pequeñas pueden incluir temas contractuales en los TdR

En evaluaciones pequeñas, los TdR pueden incluir también los detalles del proceso de licitación. En ese caso, los TdR incluirán las fechas de presentación de las ofertas, la documentación que debe adjuntarse a la oferta, el procedimiento de calificación y cualquier otro requisito puntual (por ejemplo, presentar las ofertas técnica y financiera por separado). Las evaluaciones grandes pueden incluir estos detalles en la convocatoria formal y presentar solo los detalles relativos a la evaluación en los TdR.

Cuadro 7.2: Potenciales elementos a ser incluidos en los TdR

Elemento	Comentarios
Antecedentes	Un par de párrafos son suficientes para este elemento; la mayor parte del conocimiento contextual del equipo de evaluación provendrá de la lectura de documentos relacionados con el tema. Son buenas fuentes las introducciones de las propuestas de proyectos o documentos de los llamamientos de Naciones Unidas, entre otros.
Propósito de la evaluación	¿La evaluación se usará principalmente para fines de aprendizaje o de rendición de cuentas? Si la respuesta es ambos, ¿cuál tiene prioridad? Por ejemplo, una evaluación puede haber sido pensada para dar a conocer a las instituciones donantes qué tan eficazmente se usó su dinero (rendición de cuentas), pero también podría examinar qué factores del diseño y ejecución de proyectos se llevaron a cabo en los mejores proyectos (aprendizaje). Todas las evaluaciones, estén orientadas principalmente a la rendición de cuentas o al aprendizaje, ofrecen oportunidades de aprendizaje.
Contexto de la evaluación	¿Por qué evaluar ahora? ¿Existe algún plazo que cumplir, ya sea interno o externo? ¿La evaluación está ligada a decisiones relacionadas con el ciclo de financiamiento o con la estrategia futura de la organización? Estos plazos y nexos deberían saberse si se ha involucrado a las partes interesadas primarias desde un inicio, y deben explicitarse en los TdR. Recuerden que este elemento tiene que ver con el uso; ver la <a href="#">Sección 3</a> .
Alcances de la evaluación	¿Qué sectores, área geográfica, fase de la respuesta y período de tiempo cubrirá la evaluación? ¿Se centrará en el nivel de políticas? ¿En el nivel programático? ¿En operaciones o procesos específicos? ¿Será una evaluación en tiempo real, una evaluación intermedia o una evaluación ex post? ¿Será una evaluación de una sola agencia, una evaluación propia o una evaluación conjunta? Este elemento tiene que ver con el tipo de evaluación; ver la <a href="#">Sección 4</a> .



Elemento	Comentarios
Usuarios	<p>¿Cómo utilizarán los usuarios la evaluación? La respuesta a esta pregunta debería determinar cuán largos y fáciles de leer deben ser los productos de evaluación. Ver la <a href="#">Sección 3: Pensar desde el principio y a menudo en el uso de la evaluación</a>.</p>
Marco de evaluación	<p>¿Desean que los miembros del equipo de evaluación utilicen un modelo conceptual en particular para seleccionar las metodologías de investigación y llevar a cabo su análisis – por ejemplo, el marco de medios de vida para emergencias humanitarias complejas? ¿Qué normas internacionales son relevantes para esta evaluación – por ejemplo, las normas mínimas del Proyecto Esfera o las Normas Humanitarias Esenciales? Ver la <a href="#">Sección 5: Enmarcando su evaluación</a>.</p>
Preguntas de evaluación centrales	<p>¿Cuáles son las principales preguntas que desean ver respondidas por la evaluación? ¿Qué criterios del CAD de la OCDE desean utilizar, y en qué sentido están relacionados? Ver la <a href="#">Sección 6</a> para una discusión de la selección de las preguntas de evaluación y los criterios del CAD de la OCDE.</p> <div data-bbox="470 742 1081 853" style="border: 1px solid #ccc; padding: 5px; margin: 10px 0;">  <p><b>Consejo útil:</b> Quizás prefieran limitarse a establecer los objetivos y dejar que el equipo de evaluación proponga las preguntas en el informe inicial.</p> </div> <div data-bbox="470 869 1081 981" style="border: 1px solid #ccc; padding: 5px; margin: 10px 0;">  <p><b>Consejo útil:</b> Limiten el número de preguntas para mantener enfocada la evaluación.</p> </div>
Fase inicial	<p>Los TdR deben especificar claramente qué actividades deben llevarse a cabo en la fase inicial. ¿Solo un estudio documental o también una visita inicial?</p> <p>Los TdR deben establecer también el contenido esperado del informe inicial. Ver la <a href="#">Sección 8: Fase inicial</a>.</p>
Diseños de evaluación	<p>Existen diversos diseños que pueden ser utilizados en una evaluación. Hay ciertas preguntas que pueden ser respondidas de manera óptima por un diseño en particular. Sin embargo, limitaciones logísticas y presupuestarias pueden restringir las opciones. En general, el equipo propone los diseños en el informe inicial, pero la organización que encarga la evaluación puede decidirse por un diseño en particular – para cumplir con los requisitos de la política institucional, por ejemplo. Ver la <a href="#">Sección 11: Diseños de evaluación para responder diferentes preguntas de evaluación</a>.</p>



← Cuadro 7.2: Potenciales elementos a ser incluidos en los TdR

Elemento	Comentarios
Metodologías de recopilación de información	<p>No es necesario incluir en esta sección una descripción detallada, sino más bien una indicación de sus preferencias metodológicas, si las tuvieran – por ejemplo, si desean que el equipo consulte con personas específicas tales como funcionarios estatales o personas afectadas, o que utilice alguna metodología en particular, por ejemplo una encuesta formal. Normalmente, el equipo debería desarrollar una metodología pormenorizada durante la fase inicial. Si el gestor de la evaluación decide tanto el producto como el proceso, el equipo de evaluación tenderá a desarrollar un menor sentido de apropiación y responsabilidad sobre el resultado final.</p> <p>Ver la <a href="#">Sección 10: Métodos documentales</a> y <a href="#">Sección 13: Metodologías de campo</a>.</p>
Indicadores	<p>La selección de los indicadores (si se utilizan indicadores cuantitativos) se suele dejar en manos del equipo para que los propongan en el informe inicial. Sin embargo, la organización que encarga la evaluación puede pedir que se utilicen indicadores específicos (por ejemplo, para que puedan ser comparados o consolidados con otras evaluaciones).</p>
Metodologías de análisis de información	<p>Los TdR deben indicar si existe algún requisito de utilizar un enfoque analítico en particular. Normalmente, el equipo propone el enfoque analítico que desea utilizar en el informe inicial. Ver la <a href="#">Sección 16: Análisis</a>.</p>
Cronograma	<p>Especifiquen las fechas de los principales entregables y los plazos máximos. Pregunten a las partes interesadas si hay fechas inconvenientes – por ejemplo, la temporada de lluvias o la cosecha, o bien periodos administrativos de mucho movimiento. Ver la <a href="#">Sección 9: Planificación y gestión de su evaluación</a> para una discusión de las líneas de tiempo.</p> <div data-bbox="494 1082 1108 1198" style="border: 1px solid #ccc; padding: 10px; margin-top: 10px;">  <p><b>Consejo útil:</b> Programen el tiempo suficiente para circular un informe en borrador (normalmente entre 14 y 21 días) y para la revisión del informe (otros 14 a 21 días).</p> </div>
Roles y responsabilidades	<p>Especifiquen quién es responsable de proporcionar el transporte en el trabajo de campo, concertar citas y otras tareas. Coordinen con sus colegas sobre el terreno para asegurar que otorguen el soporte necesario al equipo de evaluación.</p>



Elemento	Comentarios
Mecanismos de gestión	Especifiquen si se formarán grupos consultivos, de coordinación o referenciales y cuál será su composición y funciones. Ver la <a href="#">Sección 9: Planificación y gestión de su evaluación</a> .
Capacidades y calificaciones	¿Qué capacidades y calificaciones debe tener el equipo de evaluación para realizar esta evaluación con éxito? Estas pueden incluir: amplio conocimiento del contexto (país, sector u organización), idiomas o capacidades en metodologías o formas de comunicación específicas. Ver la <a href="#">Sección 9: Planificación y gestión de su evaluación</a> , para mayores detalles sobre las capacidades que pueden requerirse de un evaluador.
Productos	<p>Los productos habituales incluyen: un informe inicial, informes sobre el trabajo de campo (en forma de nota o presentación), el informe principal, un resumen de la evaluación y reuniones generales o cualquier otro formato considerado útil para difundir los hallazgos. Indiquen el largo y el formato si tuvieran alguna preferencia. Es conveniente asimismo establecer un cronograma de pagos escalonado, condicionado a la entrega de productos específicos, a fin de incentivar al equipo de evaluación a ponerse metas. Ver la <a href="#">Sección 17, Reportando y comunicando los hallazgos de la evaluación con un enfoque de utilización</a>.</p> <p>Si existe un estilo interno que se desea que respete el informe de evaluación, o preferencias marcadas en cuanto al formato y al estilo, indiquen cuáles son para evitar tener que corregir el informe más adelante. Especifiquen la longitud del informe (el número de palabras, que es una medida más precisa y se presta menos a malentendidos que el número de páginas) y qué anexos debería incluir, de haberlos. Quizás deseen especificar la estructura general del informe, aunque es bueno dar al equipo de evaluación cierta flexibilidad al respecto.</p>
Gestión de riesgos	Describan los riesgos y desafíos que probablemente surjan en el transcurso de la evaluación y pregunten cómo propone el equipo lidiar con ellos. Ver la <a href="#">Sección 15: Acceso restringido</a> .



← Cuadro 7.2: Potenciales elementos a ser incluidos en los TdR

Elemento	Comentarios
Presupuesto	Establezcan un presupuesto indicativo para la evaluación. Ver la <a href="#">Sección 7: Términos de referencia y presupuesto de una evaluación</a> .
Información disponible	Identifiquen las principales fuentes de información documental que estarán a disposición del equipo de evaluación. ¿Tendrán acceso a la intranet, a registros de correos electrónicos y a literatura gris (por ejemplo, evaluaciones previas del proyecto o programa)? Patricia Rogers señala que un motivo por el que las evaluaciones fracasan es que no se brinda a los equipos de evaluación "datos, información de que existen datos, evaluaciones previas, evaluaciones paralelas, cambios de política programados, cambios de personal inminentes, fechas de decisiones y reuniones clave" (comentarios enviados a Davidson, 2012b).
Evaluación de ofertas	Si no existe un documento formal de la convocatoria a licitación, quizás deseen incluir los criterios con base en los cuales se evaluarán las ofertas (qué porcentaje de la puntuación se destinará al precio, a la composición del equipo y a las metodologías de recopilación y análisis de información y otros criterios), así como la fecha límite para la recepción de ofertas. Ver la <a href="#">Sección 9: Planificación y gestión de su evaluación</a> .
Referencias clave	Posiblemente deseen adjuntar un listado de referencias clave para la evaluación. Muchas de ellas serán documentos internos, pero consulten con las partes interesadas primarias si hay otros documentos que deberían formar parte del grupo de lecturas inicial. Ver la <a href="#">Sección 10: Métodos documentales</a> .

### 7.1.3

#### Línea de tiempo de los TdR

Desarrollar los términos de referencia puede tomar apenas una semana en el caso de una organización pequeña o más de un año en el de una grande. Para concluir los términos de referencia se necesita más tiempo cuando se tienen múltiples partes interesadas repartidas en distintos países.

El IASC desarrolló un borrador de términos de referencia preaprobados para ETR interinstitucionales de emergencias humanitarias, con el fin de evitar largas negociaciones entre las partes interesadas cada vez (IASC, 2010). Este mismo formato de TdR prevé que el contrato con el equipo ganador debería suscribirse 30 días después de la fecha de publicación de los TdR y de la convocatoria a expresiones de interés. Ver la [Sección 9: Planificación y gestión de su evaluación](#) para una discusión de las líneas de tiempo.

Algunas organizaciones suelen necesitar más tiempo. Una revisión de ocho EAH con participación de los autores de esta guía reveló que el tiempo transcurrido entre la publicación de los TdR y la firma del contrato fluctuaba entre dos semanas y tres meses. La mayoría de las EAH permitían un plazo de dos meses entre la fecha de publicación de los TdR y la contratación del equipo. Para evaluaciones pequeñas, el plazo apropiado probablemente sea de un mes.

### 7.1.4

#### Extensión de los TdR

Los términos de referencia no tienen una extensión estandarizada. Una revisión de una muestra por conveniencia de 30 EAH reveló que la mayoría de los TdR constaba de entre dos y seis páginas, y tres de más de 15 páginas. Los más cortos eran de evaluaciones diseñadas para ser realizadas por un único evaluador externo o por un equipo de evaluación interno.

La longitud de los TdR de evaluaciones más complejas era mayor.



#### Consejo útil

Los TdR deben reflejar la complejidad de la tarea. Los TdR de una evaluación pequeña deben ser, en consecuencia, cortos.

El costo más elevado de una evaluación que utiliza evaluadores externos es el número de días de consultoría. Este suele ser el número de miembros del equipo y el número de días de trabajo de campo multiplicado por los honorarios y viáticos diarios. El primer paso para calcular el costo de una evaluación es estimar el número total de días requeridos. Un presupuesto puede ser dividido en tres fases: preparación, trabajo de campo y seguimiento.

## 7.2 Elaboración del presupuesto de una evaluación

### Fase inicial

La fase inicial puede constar de una reunión inicial para conversar sobre la evaluación, la lectura de documentos y de un estudio documental, una entrevista inicial y la redacción de un informe inicial.

- La reunión inicial puede tomar un par de días, incluido el tiempo de viaje, pero quizás no sea necesario que asistan todos los integrantes del equipo.
- La lectura de la documentación, las entrevistas iniciales y el informe inicial pueden tardar entre cinco y 10 días para una evaluación pequeña y varios meses o más para una evaluación grande y compleja.

En una evaluación de una sola agencia, de una sola actividad y en un solo lugar, típicamente se requerirán entre tres y siete días por evaluador para el trabajo preparatorio, dependiendo de la naturaleza de la evaluación. Al igual que la redacción del informe, una preparación apropiada a menudo requiere más tiempo de lo presupuestado. Es importante permitir cierto margen de flexibilidad.

### Trabajo de campo

Los costos del trabajo de campo – más allá del salario del personal y otros gastos relacionados con la evaluación (por ejemplo, pasajes aéreos, transporte local, traducción) – dependen de la duración del trabajo de campo así como de los honorarios y viáticos del personal y/o de los evaluadores externos.

- Como regla general, las metodologías cuantitativas demandan un trabajo de campo más extenso así como equipos de campo más numerosos y con competencias técnicas de diseño y análisis estadístico; las metodologías cualitativas requieren de un trabajo de campo más calificado.
- Una evaluación pequeña tradicionalmente destinaría una semana para actividades de observación y entrevistas de campo, media semana adicional para entrevistas informativas iniciales y otra media semana para entrevistas de seguimiento y reuniones informativas. El tiempo será más largo si se tiene previsto efectuar múltiples visitas de campo, o si se espera que el equipo elabore un informe en borrador antes de partir.
- Una evaluación pequeña tradicionalmente destinaría entre 12 y 14 días al trabajo de campo, incluido el tiempo de viaje; mientras que una evaluación grande puede requerir meses de trabajo de campo.
- Una buena regla de oro es un plazo mínimo de una semana para cada país visitado, más al menos una semana para cada lugar importante (por ejemplo, provincia o distrito).
- Si el equipo ha de realizar una consulta detallada con la población afectada, es importante programar hasta tres semanas por ubicación.
- Los métodos de encuesta cuantitativos deben ser presupuestados de acuerdo al tamaño de la muestra y al tiempo que tomará procesar y analizar la información.

**Cuadro 7.3:** Cálculo del número de días para jefes de equipo (JE) y miembros del equipo (ME)

Actividad	Días JE	Días ME
Reunión inicial, incluyendo tiempo de viaje		
Entrevistas iniciales		
Estudio documental		
Visita de campo inicial		
Elaboración de herramientas para gestionar la información		
Elaboración de plan de análisis		
Matriz de evaluación		
Informe inicial		
Reunión sobre el informe inicial		
Revisión del informe inicial		
Traslado al campo		
Sesión informativa inicial		
Reuniones en la capital		
Trabajo de campo 1		
Reunión intermedia de revisión		
Trabajo de campo 2		
Preparación de reunión informativa		
Reunión informativa		
Viaje de regreso		
Cotejo de la información		
Análisis de información		
Primer borrador del informe de evaluación		
Presentación del informe		
Revisión del primer borrador		
Revisión del segundo borrador		
Revisión final		

## Después del trabajo de campo

- Las tareas posteriores al trabajo de campo consisten principalmente en cotejar y analizar la información y redactar el informe. Nuevamente, el presupuesto programado para esta fase suele ser insuficiente. El cotejo, el análisis de la información y la redacción del informe casi siempre tardan más tiempo de lo programado, dependiendo de la complejidad del informe y de la magnitud del análisis requerido.
- Las reuniones para informar sobre el trabajo de campo tienden a durar un par de días, pero el tiempo de viaje puede comportar días adicionales. La redacción del informe puede tomar entre cinco y 20 días, y el tiempo promedio de cada ciclo de revisión es de dos a 10 días.
- Si el equipo tiene muchos miembros, una regla de oro consiste en añadir cinco días adicionales al tiempo asignado a la preparación del informe, para que el jefe del equipo pueda incorporar sus materiales.
- Una evaluación pequeña tradicionalmente destinaría entre siete y 12 días a las actividades posteriores al trabajo de campo.



### Consejo útil

No olviden sus gastos de elaboración y difusión del informe.

No esperen a que esté listo el borrador final para caer en la cuenta de que no tienen recursos para la corrección de estilo o para incluir elementos visuales (fotos, infografías, etc.). La difusión también implica un costo real (envío de portavoces a reuniones sectoriales o regionales, etc.).

En el Cuadro 7.4 encontrarán un listado de los elementos habituales de un presupuesto.

Cuadro 7.4: Elementos que generan gastos en una evaluación

Rubro presupuestario	Posibles elementos
Personal	Remuneración y viáticos del personal, viáticos del personal de la organización contraparte y de cualquier otro personal
Consultores de la evaluación	Jefe del equipo; consultores internacionales, nacionales y otros
Personal de apoyo	Remuneración y viáticos del personal administrativo, investigadores de antecedentes, intérpretes, choferes, personal de seguridad y otros
Viajes	Visas; pasajes de avión para el equipo de evaluación y el personal acompañante; transporte para asistir a reuniones informativas iniciales y reuniones para informar sobre el trabajo de campo; traslados internos del equipo y del personal acompañante
Gastos de subsistencia	Alojamiento y viáticos diarios para los consultores
Digitación de datos	Ingreso de datos y depuración de datos para eliminar las respuestas incoherentes, por ejemplo, una persona registrada como varón y embarazada
Reuniones y talleres	Alquiler del local, comidas y viáticos
Elaboración del informe	Corrección de estilo, traducción, arte, diseño gráfico, diagramación, impresión, desarrollo de medios electrónicos, distribución
Otros productos	Gastos de redacción, corrección de estilo, gráficos, etc., para la nota informativa, infografías o "10 cosas que tienes que saber sobre..."
Gastos de difusión	Gastos de viaje y honorarios para presentar los resultados a nivel nacional o regional, o en reuniones sectoriales o fuera de la zona
Gastos varios	Comunicaciones, correo y mensajería, teleconferencias, licencias y gastos judiciales, seguridad
Gastos generales	A menudo estimados como un porcentaje de otros gastos

## Notas finales

### 4 / Tipos de evaluación

1. El CAD de la OCDE ha publicado un manual de gestión de evaluaciones conjuntas (OCDE/CAD, 2006). Las evaluaciones conjuntas de intervenciones humanitarias configuran el núcleo de un capítulo de la séptima evaluación de la acción humanitaria de ALNAP (Beck y Buchanan-Smith, 2008).
2. Ver Breier (2005).

### 7 / Términos de referencia y presupuesto

3. USAID llama a los TdR “Alcance de trabajo”, que probablemente constituya una descripción más exacta.
4. Por ejemplo, la plantilla de TdR del DIFD establece que “los consultores normalmente no participarán” en la elaboración de la plantilla (DFID, 2010:1).

3/ Pensar desde el principio y a menudo en el uso de la evaluación 62 - 77

4 / Tipos de evaluación 78 - 90

5 / Enmarcando su evaluación 91 - 103

6 / Selección de las preguntas de evaluación 104 - 117

7 / Términos de referencia y presupuesto de una evaluación 118 - 131

## Notas

# Planificación y gestión de la evaluación



# 8 / Fase inicial

Esta sección incide en la fase inicial de la evaluación, y en especial en el informe inicial.



## **Definición: Fase inicial**

La fase inicial de una evaluación abarca desde la selección del equipo de evaluación hasta la aprobación del informe inicial.

En la fase inicial, el equipo intenta desarrollar una comprensión integral de la tarea de la evaluación y prepara un informe del plan de evaluación. La fase inicial tiene lugar antes del trabajo de campo; el producto es el informe inicial. Dicho informe puede ser relativamente corto, como en el caso de la evaluación del Fondo de Emergencia de Respuesta a Haití (Moriniere, 2011a), y puede ser comparado con el informe final de evaluación (Moriniere, 2011b).

La mayoría de los TdR de las EAH son fijos, es decir, no se negocian entre el gestor de la evaluación y los evaluadores. El informe inicial permite al equipo de evaluación desarrollar un plan específico y acordarlo con el gestor de la evaluación, así como plantear sus dudas y subsanar cualquier ambigüedad presente en los TdR. Por ejemplo, el trabajo implícito en los TdR puede tomar mucho más tiempo de lo presupuestado; el informe inicial es una oportunidad para que el equipo de evaluación decida qué resulta viable abarcar.

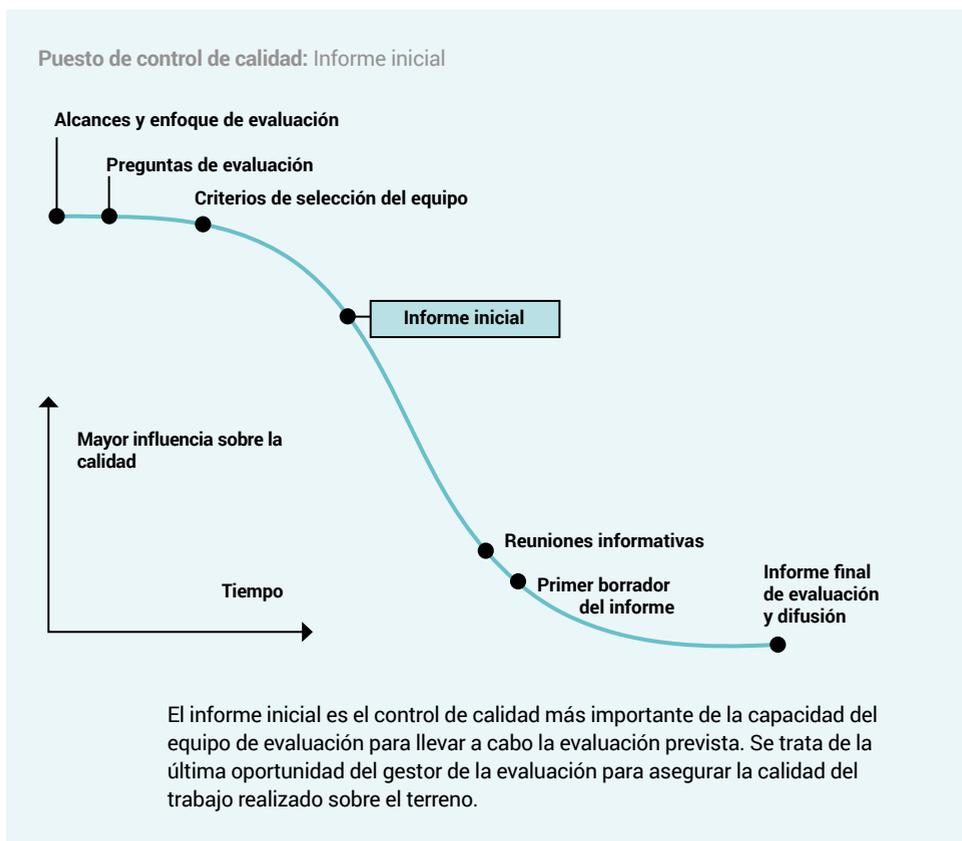


## **Consejo útil**

Elaboren un informe inicial hasta para la evaluación más pequeña. En prácticamente todos los casos vale la pena que el equipo de evaluación produzca un informe inicial, incluso si no fuera requisito hacerlo, ya que esto ayudará al equipo a planificar su trabajo y a reducir al máximo posible los malentendidos entre el equipo y el gestor de la evaluación.

## 8.1 Informe inicial y control de calidad

El informe inicial probablemente sea la herramienta más útil de la caja de herramientas de control de calidad. Además de ayudar a generar una comprensión común de la tarea de evaluación, permite asimismo al gestor de la evaluación identificar si el equipo está capacitado para llevar a cabo una evaluación de calidad. Sin embargo, si bien es cierto que un informe inicial deficiente es un buen indicativo de que el equipo de evaluación probablemente no está a la altura de la tarea, un informe de buena calidad no garantiza que lo esté. UNEG (2010) ha desarrollado una lista de verificación para los términos de referencia e informes iniciales.



## El auge del informe inicial

Los informes iniciales y las matrices de evaluación asociadas son relativamente nuevos en la EAH.

El informe inicial:

- Ayuda a que todos inicien el trabajo sobre una misma base;
- Ayuda a identificar si el alcance de la evaluación es demasiado ambicioso;
- Ayuda al gestor de la evaluación a evaluar cómo el equipo percibe y tiene previsto afrontar la evaluación;
- Le da al equipo de evaluación la oportunidad de convertir los TdR en un plan realista de común acuerdo con el gestor de la evaluación;
- Le da al equipo de evaluación la oportunidad de esclarecer los TdR y poner de manifiesto las tensiones que la organización que encarga la evaluación necesita resolver (por ejemplo, expectativas internas contradictorias con respecto a la evaluación);
- Pone de manifiesto los defectos de diseño y de las metodologías propuestas;
- Les da a las demás partes interesadas la oportunidad de recibir una declaración de intenciones clara del equipo de evaluación, que les permitirá alertar rápidamente sobre cualquier inquietud que pudieran tener con relación al enfoque propuesto;
- Ayuda a identificar si un equipo no estuviera a la altura de la tarea;
- Permite al equipo de evaluación dar a conocer con claridad lo que piensa hacer así como, lo que a veces es más importante, lo que no piensa hacer.



### Consejo útil

Para una evaluación compleja, separen la fase inicial de la fase de trabajo de campo. Para evaluaciones complejas, posiblemente convenga separar los contratos en dos partes: una para la fase inicial y la otra para la fase de trabajo de campo. A los licitadores se les invita a presentar sendas ofertas para cada fase, pero el contrato para el trabajo de campo se expide una vez aprobado el informe inicial.

## 8.2 Actividades de la fase inicial

El informe inicial debe presentar un plan de trabajo y un cronograma claros y realistas, que consideren las restricciones vigentes. El plan de trabajo debe incluir

los roles y responsabilidades de los miembros del equipo, los plazos de entrega de los informes entre los miembros del equipo y los planes de viaje detallados. El esfuerzo que se requiere para desarrollar un plan de trabajo claro depende de los alcances y la complejidad de la evaluación. La fase inicial puede incluir:

- Entrevistas iniciales con informantes clave para determinar el contexto;
- Un estudio documental, que incluya una revisión de la literatura y un estudio de información secundaria de índole cuantitativa;
- Un taller para presentar un borrador del informe inicial, que puede ser útil para validar el enfoque;
- El desarrollo, y de ser necesario la validación, de las herramientas de recolección de información;
- Una visita inicial de factibilidad al país donde se llevará a cabo el trabajo de campo, lo cual constituye una buena práctica en evaluaciones largas y complejas.



#### Consejo útil

Programen el tiempo suficiente, no solamente para que el equipo lleve a cabo una revisión documental y entrevistas iniciales, sino también entre la presentación del informe inicial y el trabajo de campo para hacer los últimos retoques.

## 8.3 La matriz de evaluación

La mayoría de los informes iniciales incluyen una matriz de evaluación, en la cual se expone un plan para responder cada una de las preguntas de evaluación. En la parte inferior encontrarán dos ejemplos, uno de una evaluación grande de un donante y otro de una evaluación de una ONG pequeña. Debido a que las matrices de evaluación son relativamente nuevas, no existe un único formato acordado, y los encabezados de las columnas pueden variar.

En su versión más simple, las matrices de evaluación pueden tener como encabezados de sus columnas las posibles metodologías; y de sus filas, las preguntas. Las intersecciones serían resaltadas de alguna manera (Todd et al., 2015).

La presente guía propone el uso de una matriz de cinco columnas:

- Preguntas de evaluación;
- Criterios de evaluación (una pregunta puede abarcar más de un criterio);
- Diseño que será utilizado para responder la pregunta;
- Metodología(s) que serán empleada(s) para responder la pregunta;
- Enfoque de muestreo que será utilizado para la pregunta.



### Consejo útil

Es conveniente que los gestores de la evaluación elaboren su propia matriz. Si bien es cierto que la elaboración de la matriz de evaluación es una tarea que le compete al equipo de evaluación, puede ser útil que el gestor prepare su propia copia interna. Dicha matriz no debe ser compartida con el equipo ganador del concurso hasta que este haya preparado su propia versión. El gestor podrá entonces comparar las dos versiones para establecer si hay vacíos importantes.

**Cuadro 8.1:** Ejemplo de una matriz de evaluación para un proyecto de transferencias de dinero en efectivo de una ONG pequeña

Pregunta	Criterios	Diseños	Metodologías	Muestreo
¿Cuál fue el impacto del programa de transferencias de dinero en efectivo en la seguridad alimentaria a nivel de hogares?	Impacto, eficacia	Diseño de “diferencias en diferencias”: comparación de los cambios a través del tiempo en la calificación de la seguridad alimentaria a nivel de hogares entre hogares receptores y no receptores*	Encuesta de hogares sobre seguridad alimentaria	Muestreo aleatorio utilizando el censo de una evaluación inicial de la seguridad alimentaria como marco muestral
		No experimental	Entrevistas cualitativas con miembros de hogares seleccionados	Selección intencional de los hogares con las puntuaciones más altas y bajas de la evaluación inicial

\* Ver la Sección 11, Diseños de evaluación para responder las preguntas de evaluación, para una descripción del diseño conocido como “diferencias en diferencias”.



Tengan presente que una pregunta puede implicar más de un diseño o metodología, y que las metodologías individuales pueden ayudar a responder más de una pregunta – no mostrada aquí.

Una vez que hayan concluido la matriz de evaluación, esta servirá de base para un cuadro de evidencia, el mismo que puede facilitar la redacción del informe, al reunir en un solo lugar toda la evidencia relacionada con una pregunta o un tema en particular (ver la [Sección 16](#) para mayores detalles sobre cuadros de evidencia).

#### **Tener presente**

El proceso de planificación siempre es más importante que el plan en sí, y el equipo debe tener la posibilidad de procesar las preguntas de evaluación en profundidad a su manera.

### Otros formatos para la matriz de evaluación

Muchas evaluaciones utilizan matrices de evaluación que solo tienen tres columnas. Estas son:

1. **Las preguntas a responder** – estas pueden ser presentadas como preguntas centrales con preguntas de evaluación subsidiarias;
2. **Cómo se formarán un juicio** – los criterios o indicadores con base en los cuales el equipo de evaluación responderá la pregunta;
3. **Las fuentes de información y las metodologías previstas** – en general hay diversas fuentes y metodologías para cada pregunta de evaluación. En ocasiones se dividen en dos columnas.

Los ejemplos que se presentan más abajo muestran extractos de las matrices de una evaluación internacional grande de un donante y de una evaluación de un solo proyecto, cada una de las cuales tiene cuatro columnas.

Una matriz de evaluación de este tipo revela la forma en que el equipo planea responder a cada pregunta, y revisarla le permitirá al gestor de la evaluación verificar si el equipo ha pasado por alto alguna fuente importante.

Algunos formatos de matrices son más complejos. Una vez lista, la matriz de evaluación puede usarse como base para elaborar un cuadro de evidencia, el mismo que reúne en un solo lugar toda la evidencia relacionada con una pregunta o un tema en particular y puede usarse para escribir el informe (ver la [Sección 16](#)). El [Cuadro 8.4](#) muestra los encabezados de las matrices de tres evaluaciones interinstitucionales.

Cuadro 8.2: Ejemplo de una matriz de evaluación para una evaluación grande

Preguntas de evaluación centrales / subpreguntas	Indicadores	Metodologías de análisis	Fuentes de información
<p><b>1. ¿Qué tan pertinente y flexible es la estrategia humanitaria danesa a la luz de los cambios ocurridos en el contexto humanitario desde 2010?</b></p> <p><b>1.1 ¿Han sido pertinentes las prioridades estratégicas, a la luz de los cambios en los desafíos humanitarios?</b></p>	<p><b>1.1a</b> Número de prioridades estratégicas cubiertas por los programas financiados por Danida</p> <p><b>1.1b</b> Coincidencia entre las prioridades estratégicas y lo que Danida y sus contrapartes consideran desafíos humanitarios clave</p> <p><b>1.1c</b> Capacidad de anticipación, adaptación e innovación de las contrapartes de Danida para afrontar la identificación y la gestión de los nuevos tipos de amenazas y las oportunidades para mitigarlos</p> <p><b>1.1d</b> Evidencia de que las estrategias financieras y nacionales de Danida son lo suficientemente flexibles como para permitir que las contrapartes se adapten a un contexto cambiante</p>	<p>Análisis de la cartera de proyectos, seguimiento de los resultados y análisis comparativo de las contrapartes para evaluar la cobertura de las prioridades estratégicas; análisis de contexto</p>	<ul style="list-style-type: none"> <li>• Mapeo de los programas de las contrapartes contra las prioridades estratégicas</li> <li>• Base de datos de los financiamientos de Danida</li> <li>• Informes de las contrapartes</li> <li>• Actas de talleres con partes interesadas sobre los desafíos humanitarios actuales</li> <li>• Revisión documental sobre el contexto humanitario internacional</li> <li>• Entrevistas con EHP y contrapartes</li> </ul>

Fuente: Mowjee et al. (2015: Anexo C)

Cuadro 8.3: Ejemplo de matriz de evaluación para la evaluación de un proyecto de una ONG

Criterios de evaluación	Preguntas centrales	Hipótesis	Fuentes de información
<b>Cobertura</b>	<b>¿Cuán apropiada fue la cobertura del programa de gestión comunitaria de la desnutrición aguda?</b>	El proyecto ha beneficiado a toda la población por igual en todas las áreas designadas.	<ul style="list-style-type: none"> <li>• Propuesta</li> <li>• Evaluaciones</li> <li>• Encuestas nacionales</li> <li>• Informes de evaluación</li> <li>• Entrevistas con informantes clave</li> </ul>

Fuente: Morán (2012: Anexo 6)

Idealmente, los criterios deben figurar inmediatamente después de las preguntas centrales.

Cuadro 8.4: Ejemplos de encabezados de matrices de evaluación

Evaluación del programa de alimentación para refugiados de larga data en Ruanda Sutter et al. (2012)	Evaluación del programa de alimentación para refugiados de larga data en Ruanda Nielsen et al. (2015).	Evaluación de la repuesta a la crisis en República Centroafricana Lawday (2015)
<ul style="list-style-type: none"> <li>• Preguntas de evaluación</li> <li>• Subpreguntas</li> <li>• Tipo de subpregunta</li> <li>• Herramientas de medición o indicadores</li> <li>• Norma o criterio (evaluación normativa)</li> <li>• Información de línea de base</li> <li>• Fuente de información</li> <li>• Diseño</li> <li>• Muestra o censo</li> <li>• Instrumento de recopilación de información</li> <li>• Análisis de información</li> </ul>	<ul style="list-style-type: none"> <li>• Tipo de pregunta (primaria o secundaria)</li> <li>• Pregunta de evaluación</li> <li>• Preguntas rectoras</li> <li>• Categoría de indicador (preliminar)</li> <li>• Metodologías</li> <li>• Fuentes de información</li> <li>• Comentarios u observaciones</li> </ul>	<ul style="list-style-type: none"> <li>• Tema</li> <li>• Preguntas</li> <li>• Subpreguntas</li> <li>• Criterios para juicios de valor, normas, lineamientos, buenas prácticas</li> <li>• Fuentes</li> <li>• Metodologías</li> <li>• Análisis</li> <li>• Fortalezas y limitaciones</li> </ul>

## 8.4 El informe inicial

El contenido del informe inicial varía, dependiendo del contexto y de la escala de la evaluación, tal como se puede apreciar en el [Cuadro 8.5](#).

Algunos elementos pueden ser incluidos en los [TdR](#) y no en el informe inicial. Esto depende de la política por la que se rija el gestor de la evaluación. El informe inicial debe dar a conocer claramente la intención del equipo de evaluación al gestor de ésta. Juntos, los TdR y el informe inicial deberán describir minuciosamente la tarea de la evaluación.

En una evaluación pequeña, el informe inicial puede incidir en solo algunos de los temas que aparecen en el cuadro pero debería contener como mínimo lo siguiente:

- La matriz de evaluación (cuya elaboración se deja en manos del equipo, como es la norma);
- El plan de trabajo (incluida la distribución del trabajo si se trata de un equipo grande);
- Las herramientas de recolección de información que se usarán.

Un informe inicial simple como este sería apropiado solamente para una evaluación muy pequeña.

**Cuadro 8.5:** Elementos indispensables y deseables en los TdR y en el informe inicial de una evaluación

<p><b>Elementos tradicionalmente incluidos en los TdR</b></p>	<ul style="list-style-type: none"> <li>• Contexto</li> <li>• Propósito (cómo se usará)</li> <li>• Objetivos</li> <li>• Criterios</li> <li>• Alcance</li> <li>• Destinatarios</li> <li>• Roles y responsabilidades</li> <li>• Hitos</li> <li>• Entregables</li> <li>• Contenido del informe inicial</li> </ul>
<p><b>Elementos que pueden ser incluidos en los TdR o el informe inicial</b></p>	<ul style="list-style-type: none"> <li>• Marco de evaluación</li> <li>• Preguntas de evaluación</li> <li>• Fuentes a ser utilizadas</li> <li>• Matriz de evaluación</li> <li>• Diseños de evaluación</li> <li>• Metodologías de recolección de información</li> <li>• Indicadores a ser medidos</li> <li>• Metodologías de análisis de información</li> <li>• Contenido del informe de evaluación</li> </ul>
<p><b>Elementos tradicionalmente incluidos en el informe inicial</b></p>	<ul style="list-style-type: none"> <li>• Plan de trabajo</li> <li>• Distribución de tareas entre los miembros del equipo</li> <li>• Herramientas de recolección de información</li> </ul>

El informe inicial debe presentar las metodologías propuestas, entre ellas un plan inicial de entrevistas prioritarias para entrevistas adicionales. Debe reconocer el rol de los grupos consultivos/consejeros, si los hubiera, y de ser posible especificarlo. En un anexo debe incluir una guía para entrevistas y una relación de los temas a tratar con los grupos focales, según proceda.

Es conveniente presentar asimismo un plan de difusión formal en el informe inicial si se espera que el equipo de evaluación participe en la difusión. Por ejemplo, si se espera que el equipo de evaluación produzca un documento académico, el informe inicial deberá identificar qué miembros del equipo serán responsables de ello.

Cuadro 8.6: El contenido de su informe inicial

Elemento	Comentarios
Antecedentes	<p>Esta sección debe resumir el contexto en el cual se está llevando a cabo la evaluación.</p> <div data-bbox="368 331 1099 443" style="border: 1px solid #ccc; padding: 5px; background-color: #f9f9f9;">  <p><b>Consejo útil</b> Incluyan una cronología, la misma que puede ser útil para exponer los antecedentes y puede ser expandida en el informe principal incluyendo la información recogida durante el trabajo de campo.</p> </div>
Acción a ser evaluada	<p>Esta sección debe demostrar que los miembros del equipo entienden qué cosas comprende la acción a ser evaluada. Puede constar de un capítulo con una descripción de la intervención e información básica recogida por el equipo durante el estudio documental, y puede incluir una teoría del cambio o algún otro modelo lógico. Puede incluir un anexo con cuadros estadísticos.</p>
Propósito de la evaluación	<p>Esta sección sintetiza el propósito y los objetivos de la evaluación desde la perspectiva del equipo, así como el uso que se dará a la misma, y explica cómo estos factores han influido en la selección de los diseños y metodologías de evaluación.</p>
Análisis de partes interesadas	<p>El análisis de partes interesadas puede ayudar al equipo de evaluación a planificar el trabajo de campo con miras a maximizar el uso de los resultados de la evaluación. Ver la <a href="#">Sección 3: Pensar desde el principio y a menudo en el uso de la evaluación</a>. El análisis de partes interesadas incluido en el informe inicial de la evaluación del clúster de seguridad alimentaria es un ejemplo (Steets et al., 2013), al igual que el cuadro del poder y los intereses incluido en el informe inicial de una evaluación interinstitucional realizada en República Centroafricana (Lawday et al., 2015: 14).</p>
Preguntas de evaluación	<p>Esta sección debe incluir cualquier formulación o reformulación de las preguntas de evaluación propuestas por el equipo. Por ejemplo, el equipo puede proponer rehacer las preguntas de evaluación para reducir el total a un número manejable, o bien para enfocarse en los temas centrales. Estos se encuentran resumidos en la matriz de evaluación. Ver la <a href="#">Sección 6: Selección de las preguntas de evaluación</a></p>
Diseño de evaluación	<p>Si bien el diseño se encuentra resumido en la matriz de evaluación, esta sección presenta el razonamiento detrás de la selección del diseño o diseños. En la actualidad pocas EAH identifican sus diseños. La evaluación de la asistencia alimentaria para los refugiados de larga data en Bangladesh (Nielsen, 2012: 5) es una excepción, que se identifica a sí misma como un diseño ex post-facto con grupo de control no equivalente. Ver la <a href="#">Sección 11: Diseños de evaluación para responder las preguntas de evaluación</a>.</p>
Metodologías	<p>Las metodologías se encuentran resumidas en la matriz de evaluación, pero en esta sección se presentan más detalladamente las metodologías que el equipo tiene previsto utilizar para recolectar y analizar la información con el objeto de responder las preguntas de evaluación. Ver la <a href="#">Sección 10: Métodos documentales</a> y <a href="#">Sección 13: Metodologías de campo</a>. El detalle de cada metodología debe indicar la estrategia de muestreo prevista y el enfoque analítico propuesto. Debe especificar claramente las limitaciones de las metodologías de recopilación de información propuestas, entre ellas la estrategia de muestreo, así como cualquier limitación a la disponibilidad de recursos. Ver la <a href="#">Sección 12: Muestreo</a>.</p>



Cuadro 8.6: El contenido de su informe inicial

Elemento	Comentarios
Herramientas de recolección y análisis de información	Estas suelen adjuntarse como anexos del informe inicial e incluyen guías para las entrevistas, formularios de encuestas, rúbricas u otros instrumentos de recolección de información a ser utilizados. Deben presentarse también los instrumentos de análisis de información, incluyendo cualquier listado de códigos inicial. Ver la <a href="#">Sección 16: Análisis</a> .
Matriz de evaluación	La matriz de evaluación revela la forma en que los evaluadores planean responder a cada una de las preguntas de evaluación. Debe reflejar los diseños y metodologías propuestos en el informe inicial. Puede presentarse en forma de anexo. Mientras que el cuerpo del informe detalla los diseños y metodologías, la matriz revela cuáles tiene previsto utilizar el equipo para cada pregunta.
Plan de trabajo detallado	<p>Esta sección especifica los lugares que los miembros del equipo tienen previsto visitar y cuándo, así como los días propuestos para las visitas a la oficina central. Debe establecer asimismo las responsabilidades de cada miembro del equipo. Ver la <a href="#">Sección 7: Términos de referencia y presupuesto de una evaluación</a>, así como el análisis de líneas de tiempo en la <a href="#">Sección 9</a>.</p> <div data-bbox="349 683 1084 815" style="border: 1px solid #ccc; padding: 10px; background-color: #f9f9f9;"> <p> <b>Consejo útil</b> Eviten los planes detallados en los entornos inseguros. La razón es que revelar los planes de viaje específicos con anticipación puede implicar un riesgo de seguridad, y a menudo se requiere flexibilidad para decidir a última hora en dichos entornos.</p> </div>
Estructura del informe central y formato de otros productos	Esta sección suele presentarse en forma de índice y puede incluir detalles de otros productos de evaluación, por ejemplo, un esbozo todavía preliminar de los talleres de difusión y cualquier otro producto de evaluación. Puede incluir asimismo un plan de difusión detallado. Ver la <a href="#">Sección 17: Informes y comunicación de los resultados de evaluación con un enfoque de utilización</a> .
Relación de posibles entrevistados	Esta sección contiene una relación preliminar de las personas que el equipo planea entrevistar, o al menos de los tipos de personas a ser entrevistadas.
Preguntas y temas pendientes	Esta sección constituye una oportunidad para poner de manifiesto las ambigüedades, áreas de preocupación o contradicciones que el equipo de evaluación quisiera que la organización que encarga la evaluación examine y aclare antes de pasar a la siguiente etapa.
Riesgos y medidas de mitigación	El equipo identifica qué riesgos avizora y qué medidas piensa tomar para atenuarlos.
Consideraciones éticas	En esta sección el equipo explica cómo afrontará los aspectos éticos de la evaluación. Ver la <a href="#">Sección 14: Involucrando a la población afectada en su evaluación</a> .

## Evaluando el informe inicial

El informe inicial permite que el gestor de la evaluación identifique la percepción del equipo de la tarea de evaluación y el enfoque que piensa utilizar para acometerla. Permite, además, que las partes interesadas en terreno tomen conocimiento de cómo planea ejecutar el trabajo el equipo y así identifiquen cualquier problema o desafío pertinente.

El informe permite asimismo que el gestor de la evaluación procese cualquier diferencia de percepción y enfoque con el equipo antes de que se conviertan en problemas mayores. El informe debe demostrar la percepción del equipo del contexto de la crisis humanitaria, el contexto de la respuesta y de las acciones a ser evaluadas, el propósito y la intención de la evaluación y las inquietudes de las partes interesadas.

El informe inicial podría ser evaluado con base en el [Cuadro 8.6](#) que se mostró anteriormente, considerando lo siguiente:

- ¿Debería contener este informe inicial una sección así?
- ¿La sección revela claramente lo que se tiene programado hacer?
- ¿Es adecuado y apropiado el contenido de la sección?

Las respuestas a estas preguntas sacarán a la luz las áreas que el gestor de la evaluación necesita tratar nuevamente con el equipo de evaluación.



### Consejo útil

Especifiquen claramente qué es lo que desean. Si desean que el equipo incida en determinados aspectos en el informe inicial, asegúrense de especificar cuáles son dichos aspectos en los TdR.

# 9 / Planificación y gestión de su evaluación

Esta sección describe diversos mecanismos de gestión y gobernabilidad para una evaluación. Incide asimismo en las líneas de tiempo, la contratación del equipo y los desafíos para el liderazgo de una evaluación.

La escala y la complejidad de los mecanismos de gestión y gobernabilidad de una evaluación deben ser proporcionales a la escala y la complejidad de la evaluación. Las evaluaciones conjuntas de gran tamaño tienden a requerir los mecanismos de gobernabilidad más complejos.

## 9.1 La decisión de evaluar

La tarea del gestor de la evaluación empieza cuando se toma la decisión de llevar a cabo o considerar llevar a cabo una evaluación. Ver la [Sección 2, La decisión de realizar una evaluación](#).

Esta decisión puede tomarse en la etapa de diseño de un proyecto o durante la ejecución del mismo, o puede ser parte de la política de evaluación de la organización. Una política institucional de evaluación puede establecer la obligatoriedad de evaluar los programas de ciertas dimensiones o de cierto nivel de complejidad.

En otros casos, el director de programas, la alta dirección o el gestor de la evaluación pueden proponer una evaluación, ya sea de acuerdo con la [política institucional de evaluación](#) más amplia o decidiendo caso por caso.

## El sesgo positivo de la evaluación

Los directores de programas tienden a favorecer las evaluaciones de programas considerados exitosos. Puede haber cierta resistencia a evaluar aquellos programas que han sido considerados ineficaces, ya que los directores pueden temer que los resultados impacten negativamente en su carrera profesional. Esto introduce un sesgo hacia los resultados positivos en la actividad de evaluación.

Por otro lado, los gestores de evaluaciones que al mismo tiempo lideran el aprendizaje corporativo más amplio pueden preferir evaluar programas considerados relativamente infructuosos debido al potencial de estos para extraer enseñanzas. En dichos casos, los gestores de las evaluaciones pueden tener que realizar una labor de convencimiento para obtener respaldo para la evaluación.



### Consejo útil

Una visita de campo del gestor de la evaluación antes de la misma puede ser útil para dar los últimos retoques a los TdR y aplacar cualquier duda sobre la evaluación que pudieran tener las partes interesadas sobre el terreno (ver la discusión sobre análisis de evaluabilidad).

## 9.2 Grupos consultivos

Los mejores mecanismos de gestión de una evaluación son aquellos en que las partes interesadas primarias permanecen involucradas en las decisiones. Para ello se forman a menudo grupos consultivos. La tipología de grupos consultivos más comunes se describe más abajo. Mientras que una evaluación de gran tamaño puede utilizar todos estos grupos consultivos, una evaluación pequeña puede utilizar una estructura de gestión mucho más simple, con un grupo consultivo integrado por unos cuantos colegas de evaluaciones de organizaciones similares. Estos ejercerán las funciones de coordinación y referencia de pares, y el gestor de la evaluación, las de dirección y gestión.



### Consejo útil

En vez de crear un grupo consultivo para una sola evaluación, consideren establecer un grupo de colegas de evaluaciones de organizaciones similares para que funjan de grupo consultivo permanente para las respectivas evaluaciones de cada organización. También puede ser una oportunidad para forjar alianzas con instituciones académicas.

## Grupo de coordinación



### **Definición: Grupo de coordinación**

Grupo establecido para encauzar una evaluación a través de etapas centrales como la elaboración de los TdR, la redacción del informe inicial y la redacción del informe final.

En una evaluación de grandes dimensiones, y especialmente en una evaluación conjunta, el grupo de coordinación tradicionalmente está formado por los gestores de evaluaciones de varias organizaciones. Un liderazgo fuerte resulta indispensable para que el grupo trabaje con eficacia y hasta el máximo de sus potencialidades, especialmente si la membresía es diversa. En una evaluación pequeña, el comité de coordinación de la evaluación puede estar formado por el gestor de la evaluación y uno o dos colegas. El grupo de coordinación se reúne solamente en las etapas más importantes de la evaluación.

La evaluación de la respuesta a la emergencia de Etiopía en 2002-2003 fue dirigida por un comité de coordinación formado por representantes del gobierno, donantes, el sistema de las Naciones Unidas y las ONG participantes (Simpkin et al., 2004: 4). Para la evaluación del programa de transversalización del enfoque de edad, género y diversidad de ACNUR, el comité de coordinación (denominado “comité ejecutivo”), integrado por personal de ACNUR, gobiernos y ONG, se reunió tres veces en Ginebra. El comité examinó y validó las metodologías de evaluación, revisó los informes interinos y ofreció retroalimentación sobre las conclusiones y recomendaciones (Thomas y Beck, 2010: 10).

Los comités de coordinación también se usan en el caso de evaluaciones pequeñas. La evaluación conjunta de CARE–Save the Children de la respuesta al terremoto de Haití contó con un comité de coordinación compuesto por un representante de cada organización (O’Hagan et al., 2011: 37). En este caso, el comité de coordinación seleccionó los lugares para las visitas de campo (ibid.: 13). Si bien son más habituales en evaluaciones conjuntas, los comités de coordinación también pueden utilizarse en evaluaciones de una sola institución. La evaluación del Consejo Noruego para Refugiados (NRC) de su programa de información, orientación y asistencia legal en Afganistán y Pakistán contó con un comité de coordinación integrado por cinco personas: la Directora Regional y cuatro miembros del personal de la oficina central (Pierce, 2009: 10).



## Comité de gestión

**Definición: Comité de gestión**

Grupo que gestiona la evaluación diariamente, entre otras cosas la redacción de los TdR, la contratación y gestión del equipo de evaluación y la gestión de la revisión y culminación del informe de evaluación.

Si el comité de coordinación tiene muchos miembros, es recomendable establecer un comité de gestión más reducido, que pueda tomar decisiones rápidamente cuando proceda, sin necesidad de convocar a una reunión de todo el grupo de coordinación. Tradicionalmente, el comité de gestión se reúne más frecuentemente que el grupo de coordinación. Las evaluaciones de menor tamaño pueden recurrir a un comité de gestión en vez de a un comité de coordinación. Una evaluación pequeña puede ser ejecutada por el gestor de la evaluación, sin necesidad de un comité o grupo de coordinación.

La ETR de tres meses de duración de la respuesta al terremoto de Haití en 2010 contó con un grupo de gestión integrado por representantes de OCHA, UNICEF y una ONG (Grünewald et al., 2010: 19). En la ETR de UNICEF de la respuesta a la crisis de Malí, los TdR exigían que el comité de gestión se reuniera semanalmente para evaluar los avances (Leonardi y Arqués, 2013: 90). La estructura del informe de evaluación fue acordada conjuntamente por el equipo de evaluación y el comité de gestión (ibid: 24).

En evaluaciones pequeñas, el gestor de la evaluación puede asumir todas estas tareas.

## Grupo referencial

**Definición: Grupo referencial**

Grupo formado por partes interesadas primarias familiarizadas con el entorno local, que pueden prestar asesoría sobre cuestiones prácticas relacionadas con la evaluación y sobre la viabilidad de las recomendaciones.

Mientras que los términos “comité de coordinación” y “comité de gestión” se usan de manera bastante consistente, lo mismo no puede ser afirmado de los términos “grupo referencial” y “revisores de pares”. El grupo referencial de una organización es el grupo de revisión de pares de otra. Los estándares de evaluación de UNEG hacen referencia a “un grupo de revisión de pares o grupo referencial integrado por expertos externos” (UNEG, 2005: 15). El PMA hace una distinción entre grupo

referencial interno y revisores de pares externos. La evaluación estratégica de los planes de contingencia del PMA contó con un grupo referencial interno de seis personas y un grupo de revisión de pares externo de tres miembros (Ressler et al., 2009: ii). El informe de evaluación de la participación danesa en Somalia y alrededores (Gardner et al., 2012: 18) fue sometido a revisión tras recibir los comentarios del grupo referencial.

Establecer un grupo referencial es una buena manera de involucrar a las partes interesadas primarias cuando una evaluación ha sido encargada por la sede central de la organización. La evaluación de UNICEF del programa de educación en Timor Oriental contó con un grupo referencial de partes interesadas clave identificadas por la oficina de país (Tolani-Brown et al., 2010: 93)



#### **Ejemplo de buena práctica: Capitalizando la participación de la oficina de país en una evaluación**

Para hacer de la evaluación el ejercicio más útil posible a nivel de la oficina de país, la Oficina de Evaluación de UNICEF buscó involucrar a la oficina de Timor Oriental en todas las etapas de la evaluación de su programa de educación, creando un grupo de referencia local. La coordinadora de la evaluación viajó a Timor Oriental e involucró al oficial de seguimiento y evaluación de la oficina en las discusiones sobre la estrategia de muestreo y el diseño de la encuesta, apoyando asimismo con la calidad de la evaluación con miras a robustecerla. Este componente de desarrollo de capacidades sirvió de incentivo para que el personal de SyE apoyara activamente el programa de educación y apostara por el éxito del mismo. El resultado fue una evaluación de alta calidad y pertinencia, que generó una respuesta de la dirección y un proceso de seguimiento fuerte dirigido por la oficina país. Como consecuencia de este enfoque independiente y sin embargo colaborativo, la institución donante comprometió un sustancial segundo tramo de financiamiento multianual para el programa. Este proceso provocó asimismo un cambio en la cultura institucional de la oficina país, que se volvió mucho más partidaria de la evaluación.

Fuente: UNICEF (2010)

## **Grupo de revisión de pares**



#### **Definición: Grupo de revisión de pares**

Grupo que asesora sobre aspectos relacionados con la calidad, generalmente formado por evaluadores y otros especialistas seleccionados por su conocimiento de evaluación, de la región o del tipo de intervención bajo evaluación.

Un grupo de revisión de pares no tiene por qué tener muchos miembros. Puede asesorar con el análisis contextual, el diseño y las metodologías y comentar el informe en borrador. Puede ser especialmente provechoso en evaluaciones que son coordinadas por un gerente general en vez de un evaluador y cuando el equipo de evaluación no está familiarizado con la región. Para la revisión del Comité de Emergencia para Desastres (DEC) de la respuesta a la crisis de Kosovo en 1999 se creó un equipo de revisión de pares consistente de expertos regionales y humanitarios que aleccionaron al equipo de evaluación antes de viajar y posteriormente discutieron el informe preliminar con ellos (Wiles et al., 2000: 4). La evaluación de la cooperación de la FAO en Somalia anotó que los dos revisores de pares habían suministrado “insumos y retroalimentación constructivos y reveladores” (Buchanan-Smith et al., 2013: v). La evaluación fue sometida asimismo a un proceso de revisión de pares al interior del departamento de evaluación (ibid: 4).

El número de revisores de pares varía. La evaluación de Oxfam GB de su programa de seguridad alimentaria y medios de vida urbanos en Nairobi (Macaulan y Phelps, 2012) contó con un revisor de pares externo. La evaluación del DEC de la respuesta a las inundaciones de Mozambique en el año 2000 contó con tres revisores de pares (Cosgrave et al., 2001, pág. 1), a quienes el jefe del equipo hizo llegar el borrador del informe de evaluación al igual que a los demás miembros del equipo (ibid. p. 107). Los revisores de pares revisaron también la versión final.



#### **Consejo útil**

Mantengan a los grupos de revisores de pares y de referencia separados, de modo tal que cada uno pueda enfocarse en la tarea que le ha sido asignada y evitar las distracciones. Aunque sus responsabilidades parecieran coincidir, funcionan mejor por separado.



#### **Consejo útil**

Incluyan una partida presupuestaria en el presupuesto de la evaluación para pagar los honorarios de los revisores de pares que estén trabajando de manera independiente, en vez de como parte de un emplee remunerado.

Las evaluaciones conjuntas de gran tamaño se beneficiarán de contar con toda la gama de grupos consultivos, pero para las evaluaciones más pequeñas los mecanismos más simples resultan más apropiados. Hasta la más pequeña de las evaluaciones puede beneficiarse de contar con un grupo referencial de partes interesadas e incluso con un único revisor de pares.

## 9.3 Línea de tiempo de la evaluación

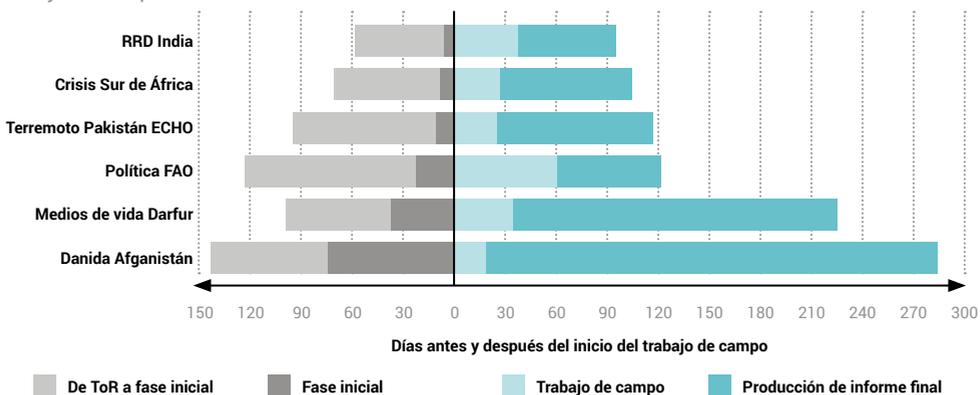
La duración exacta de una evaluación dependerá del contexto. El cuadro que presentamos a continuación ofrece algunas estimaciones.

Cuadro 9.1: Línea de tiempo estimada de una evaluación

Actividad	
<b>Desarrollo de los TdR</b>	Entre una semana y más de un año, dependiendo de la magnitud de la consulta y del nivel de acuerdo requerido. El IASC ha desarrollado una plantilla de TdR que puede ser rápidamente adaptada (IASC, 2010). Este enfoque es especialmente apropiado cuando se tiene prevista una serie de evaluaciones similares, o para una organización enfocada en un tema en particular (por ejemplo, un enfoque o parámetros institucionales centrados en la infancia).
<b>Publicación de los TdR hasta la decisión de adjudicar el contrato</b>	Esta etapa puede ser bastante rápida en el caso de una evaluación interna si ya se cuenta con el personal disponible. Toma más tiempo cuando se contrata a consultores externos. Graig (2013) sugiere que entre tres y cuatro semanas es un plazo suficiente para que los evaluadores presenten una expresión de interés, pero que puede requerirse más tiempo si se trata de un pedido de expresiones de interés amplio. Es más común que esta etapa más la adjudicación del contrato duren dos meses.
<b>Formalización del contrato</b>	Entre un día y varios meses, dependiendo de la organización.
<b>Fase inicial</b>	Entre una semana y varios meses o más. Las fases iniciales son más largas cuando se requiere una visita inicial y/o una revisión documental extensa.
<b>Trabajo de campo</b>	Entre una semana (para evaluaciones más pequeñas) y varios meses, dependiendo de la escala y la complejidad de la evaluación. Entre dos y cuatro semanas es el plazo más común para EAH.
<b>Análisis de la información y redacción del informe</b>	Entre una semana y varios meses, dependiendo del nivel de complejidad. Un análisis de la información de encuestas puede tomar varios meses si es necesario traducir o codificar la información.
<b>Ciclo de revisión</b>	Cada ciclo de revisión consta de un período para comentarios, tradicionalmente de dos semanas en el caso de una evaluación pequeña y hasta cuatro semanas en una evaluación más compleja en la que las organizaciones tienen que consultar con su personal en terreno. A ello le siguen una o dos semanas para la revisión del borrador del informe. Virtualmente todos los planes de evaluación incluyen un ciclo de revisión pero pueden requerirse muchos más, especialmente cuando los resultados de la evaluación no son necesariamente bienvenidos o si existe la necesidad de generar un consenso.
<b>Difusión</b>	Esta fase puede implicar la realización de talleres o visitas de campo para difundir los resultados, generalmente dentro del plazo de un mes a partir de la fecha de publicación del informe.

El gráfico que aparece a continuación muestra las líneas de tiempo de ocho evaluaciones de distintos tipos en las que participaron los autores. En la mayoría de los casos se trata de evaluaciones de tamaño mediano. Las evaluaciones más pequeñas fueron más rápidas que las más largas.

**Figura 9.1:** Líneas de tiempo de evaluaciones con los plazos en días, antes y después del inicio del trabajo de campo



Hubo razones específicas para que las distintas fases tomaran tanto tiempo en cada caso. El objetivo de la ilustración es por lo tanto demostrar la amplia variedad y no presentar normas.

## 9.4 Evaluación interna y externa

Una evaluación interna es aquella realizada por miembros del personal de la organización que está siendo evaluada. Algunas organizaciones utilizan el término interno y externo no para referirse a la organización sino al programa, por ejemplo, califican a una evaluación de externa cuando es realizada por personal de la organización que no está vinculado al programa. Una evaluación externa es realizada por un equipo de evaluación que es externo a la organización bajo evaluación.

Algunas organizaciones tienen departamentos de evaluación cuyo personal tiene una trayectoria profesional específica y responde directamente al comité ejecutivo, y no a los oficiales de programas. Dichas organizaciones argumentarían que en tales circunstancias incluso una evaluación interna puede ser independiente.

En algunos casos una evaluación puede ser ejecutada por un equipo compuesto por consultores externos y personal interno, en cuyo caso se trata de una evaluación mixta. Si la persona que encabeza la evaluación es externa, se considera una evaluación independiente. La combinación de un jefe externo con personal interno puede facilitar el aprendizaje.

La decisión de utilizar un equipo de evaluación externo, interno o mixto dependerá del propósito de la evaluación (ver la [Sección 2, La decisión de realizar una evaluación](#)). Si el fin principal de la evaluación es la rendición de cuentas, los evaluadores deben ser externos, ya que es posible que evaluadores internos no sean percibidos como suficientemente independientes. Dependiendo del contexto institucional, puede ser deseable incluir uno o más miembros del personal como parte del equipo – quizás del departamento de evaluación de la organización, si lo hubiera, o, de ser apropiado, de los departamentos operativos. Dichos integrantes nunca deben ser mayoría ni asumir la jefatura de la evaluación, ya que incluir personal interno en el equipo puede sembrar dudas con relación a la independencia de la evaluación.

Si el fin principal de la evaluación es el aprendizaje, el equipo de evaluación debe consistir ya sea mayoritariamente de los miembros del personal interno de los cuales se espera que aprendan o de un equipo de evaluación externo cuya función principal sea facilitar el aprendizaje del personal. Los equipos mixtos pueden ayudar a completar el círculo de aprendizaje, ya que el personal de la organización podrá aportar ideas sobre cómo poner en práctica de manera óptima su aprendizaje institucional. Por ejemplo, pueden saber qué partes interesadas clave deberían ser involucradas y qué áreas de la organización están abiertas al cambio (Riccardo Polastro, comunicación personal, junio de 2013).



#### **Pregunta clave**

Los consultores externos son considerados más independientes que el personal de la organización frente a las presiones de la dirección para presentar un programa bajo una luz favorable. Esto es cierto por regla general pero, si un evaluador o una consultora especializada en evaluación tiene interés en seguir trabajando para un cliente, ¿qué impacto puede tener esto en su independencia?

Los pros y contras de usar evaluadores internos y externos se presentan en el [Cuadro 9.2](#) de la siguiente página.

Cuadro 9.2: Ventajas y desventajas de utilizar evaluadores internos y externos

Evaluadores internos		Evaluadores externos	
+	Los beneficios que derivan del aprendizaje y la reflexión durante el proceso de evaluación permanecen dentro de la organización.	+	Son más objetivos.
+	Conocen la organización y su cultura institucional.	+	Es menos probable que presenten sesgos institucionales.
+	Son conocidos por el personal.	+	Introducen una perspectiva fresca.
+	Pueden parecer menos intimidantes y más confiables.	+	Pueden tener mayor experiencia en la cual apoyarse.
+	Los hallazgos y recomendaciones pueden ser más apropiados para la organización.	+	Posiblemente puedan necesitar más tiempo para la evaluación.
+	Las recomendaciones a menudo tienen mayores probabilidades de ser puestas en práctica.	+	Pueden servir como expertos o facilitadores externos.
+	Son menos costosos.	+	No forman parte de la estructura de poder de la organización.
+	Desarrolla capacidades de evaluación internas y en general contribuye a la capacidad del programa.	+	Pueden aportar recursos adicionales.
-	Su objetividad puede ser cuestionada.	+	Es probable que estén capacitados y tengan experiencia en evaluación.
-	La estructura institucional puede restringir la participación.	+	Son considerados expertos.
-	Compromisos de trabajo pueden restringir la participación.	-	Es posible que no conozcan la organización.
-	Su motivación puede ser cuestionada.	-	Posiblemente no conozcan las restricciones que afectarán la aceptación de las recomendaciones.
-	Pueden aceptar con demasiada facilidad los supuestos de la organización.	-	Los beneficios que deriven de la reflexión y el aprendizaje durante el proceso de evaluación no permanecerán dentro de la organización.
-	Es posible que no estén capacitados en metodologías de evaluación.	-	Pueden ser percibidos como adversarios.
-	Pueden mermar la credibilidad de la evaluación fuera de la organización.	-	Son más costosos.
-	Les puede ser difícil evitar los sesgos.	-	Su contratación puede requerir extensas negociaciones contractuales.
-	Es posible que carezcan de conocimientos y experiencia técnica especializada.	-	Es menos probable que realicen un seguimiento de las recomendaciones.
		-	Pueden no estar familiarizados con el entorno.
		-	Pueden estar influenciados por la necesidad de asegurar contratos futuros y por lo tanto ser menos independientes de lo que aparentan.

Recopilación propia de los autores Ver Gosling y Edwards (2003), así como el documento de la serie *Shortcuts* de CRS y la Cruz Roja Americana, *Preparing for an Evaluation* (2008).

## Aprendizaje de las partes interesadas en una evaluación

Realizar una evaluación ofrece excelentes oportunidades de aprendizaje al equipo de evaluación. Si el aprendizaje de las partes interesadas es prioritario, una evaluación externa puede no ser el mejor enfoque. Las evaluaciones internas permiten al personal de la organización capturar el aprendizaje – aumentando las probabilidades de que las enseñanzas sean puestas en práctica. Las evaluaciones mixtas, posiblemente encabezadas por un evaluador externo experimentado e integradas por algunos miembros del personal de la organización, representan una buena oportunidad de aprendizaje, al mismo tiempo que mantienen la calidad y la credibilidad de la evaluación.

Un enfoque adicional es el modelo de evaluación de pares. Este puede ser implementado de una de dos maneras:

- Otras organizaciones contrapartes participan en la evaluación de una contraparte; o
- Programas de distintos países participan en la evaluación de un programa similar en otro país.

Este enfoque asegura que quienes aprenden de la evaluación sean personas en condiciones de poner en práctica lo aprendido en otros contextos. Nuevamente, emplear a un evaluador externo experimentado contribuye a garantizar la calidad y la credibilidad de la evaluación.

## 9.5 Contratación de evaluadores

Los evaluadores externos tienen que ser contratados. Los coordinadores de evaluaciones disponen de una serie de opciones para hacerlo:

- Emitir una convocatoria formal, especificando los detalles de la evaluación en el anexo con los TdR o como parte de la convocatoria a presentar propuestas;
- Emitir TdR que contengan los detalles del contrato;
- Expedir una carta de intención manifestando el interés de la organización en mandar hacer una evaluación.

Algunas organizaciones emplean el término “carta de intención” para referirse a una expresión de interés. En esta publicación, utilizamos el término para referirnos al documento que presentan los evaluadores a una organización para dar a conocer su interés en llevar a cabo una evaluación.

Davidson (2012) y Graig (2011; 2013) son partidarios del uso de un proceso de expresiones de interés para seleccionar al evaluador o identificar a los posibles consultores. El uso de expresiones de interés para seleccionar a los evaluadores es especialmente indicado para evaluaciones pequeñas. Davidson (2010b) sugiere preguntar a quienes presentan una expresión de interés lo siguiente:

- ¿Qué o quién es la identidad, persona o equipo que desea presentar una oferta para hacer la evaluación? (hasta una página)
- ¿Por qué les interesa esta evaluación? (hasta media página)
- ¿Qué piensan que pueden hacer? ¿Qué conocimientos especializados, experiencia y capacidad poseen? (hasta una página)
- ¿Quiénes son – una empresa consultora especializada en evaluación, un evaluador o un equipo de evaluación? ¿Cuáles son sus valores, prácticas y áreas de expertise y especialización distintivas? (hasta media página)
- ¿Cuáles son las tarifas diarias que cobran los supuestos miembros del equipo y qué otros gastos generales o imprevistos requerirían?
- De ser posible, adjunten dos o tres resúmenes ejecutivos de informes de evaluación recientes que haya(n) realizado y/o dirigido el evaluador o evaluadores responsable(s).

Davidson sugiere, además, no pedir a los evaluadores su curriculum vitae (CV) u otros documentos adicionales en la etapa de preselección. Los CV pueden ser apropiados cuando un contrato se va a basar exclusivamente en expresiones de interés, como podría ser el caso en una evaluación pequeña.

Davidson añade que la convocatoria habitual termina con ofertas pormenorizadas, llenas de detalles triviales, con escasa información que las distinga unas de otras. La autora sugiere que los evaluadores deben ser seleccionados de acuerdo no a su plan de evaluación sino a su capacidad para hacer frente a los desafíos más importantes. He aquí algunas áreas de interrogación posibles:

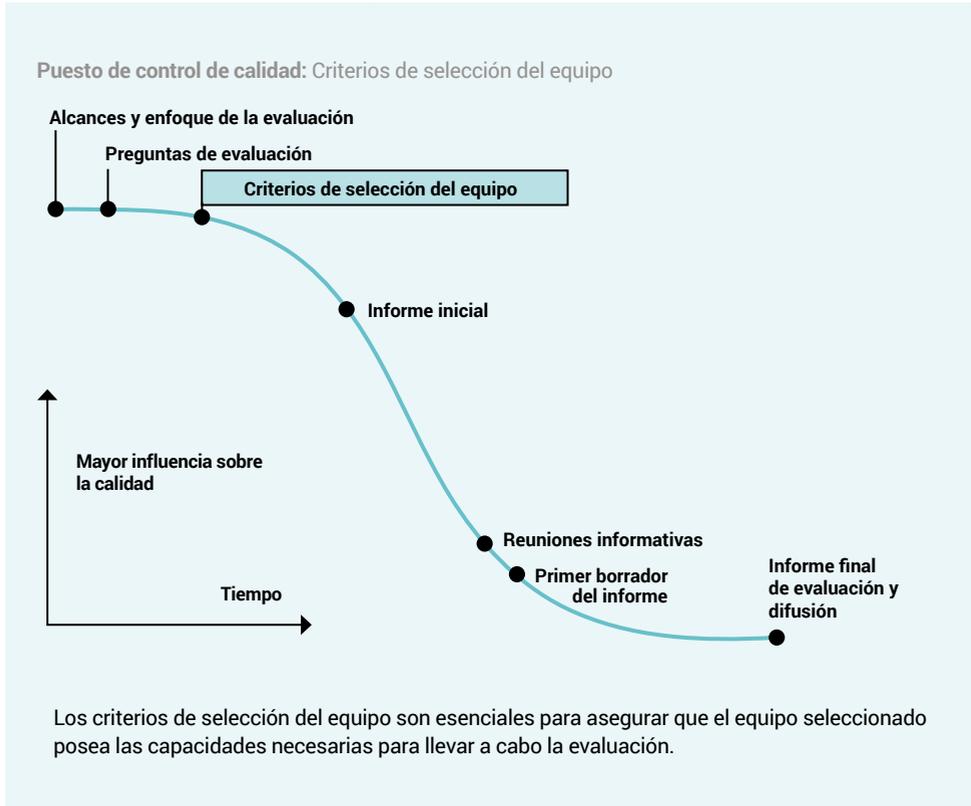
- Si se trata de una evaluación, en contraposición a un proyecto de investigación sobre el tema, pregunten cómo definen la diferencia. En su opinión, ¿qué conocimientos y destrezas necesitan para postular a una evaluación y no a un proyecto de investigación?
- ¿Cómo harían para contestar una pregunta como esta: “¿Qué tan valiosos son los efectos para [un sector de beneficiarios específico]?”
- ¿Cómo gestionarían diferencias de opinión sobre los resultados más significativos y cómo sabrían reconocer la calidad? ¿Cómo las aplicarían adecuadamente para extraer conclusiones?
- ¿Cómo manejarían los retos previstos que enfrentará la evaluación? Soliciten ejemplos de cómo lo han hecho en el pasado.

- ¿Cómo se describirían a sí mismos en su calidad de evaluadores profesionales? ¿Cuál es su “imagen de marca”? ¿Cuál es su “enfoque insignia”, por el que son más conocidos? ¿Qué tipos de proyectos sacan mayor partido de sus fortalezas? ¿Qué tipos de proyectos evitan por encontrarse fuera de sus áreas de fortaleza? Sondeen su expertise y su enfoque evaluativo; no los dejen evadir la pregunta con una respuesta sobre su expertise en contenidos.
- Pídanles tres resúmenes ejecutivos de evaluaciones en las que hayan trabajado recientemente. Estos son muy reveladores de las capacidades de los evaluadores para redactar, ir al grano, expresarse de manera concisa y responder preguntas importantes desde un punto de vista evaluativo (y no solo relacionarlas con datos), así como cuán bien entienden realmente a los usuarios previstos y lo que tendría sentido para ellos (Davidson, 2012b).

Davidson (2010c) sugiere que estas preguntas podrían ser respondidas por escrito o por medio de una presentación. Estas preguntas, u otras similares, podrían formar la base para una entrevista (personal o electrónica) con el jefe del equipo. Según el enfoque de Davidson, esta entrevista cubriría buena parte de los aspectos técnicos.

## Selección del equipo externo adecuado

La definición de los criterios de selección para la oferta ganadora tendrá una fuerte influencia en la calidad de la evaluación. Estos pueden ser incluidos en los TdR o en una convocatoria aparte.



Los criterios de selección deben incluir:

- Conocimiento necesario sobre el contexto;
- Capacidades de comunicación necesarias;
- Capacidades para llevar a cabo un trabajo de campo de calidad;
- Capacidades analíticas para convertir la evidencia recogida en el campo en respuestas convincentes a las preguntas de evaluación.

<b>Criterios cualitativos para declarar ganadora una oferta</b>	<b>Puntuación máxima</b>
<b>Comprensión de los TdR y del objetivo de los servicios a prestar</b>	10
<b>Metodologías de estructuración y de recolección y análisis de información</b>	40
<b>Organización de tareas y del equipo, cronograma de trabajo</b>	10
<b>Idoneidad del equipo con base en los conocimientos y experiencia propuestos</b>	40
<b>Puntuación técnica total</b>	100
<b>Factor precio: oferta más económica/esta oferta</b>	
<b>Puntuación total: Puntuación técnica total multiplicado por el factor precio</b>	

Basado en Maunder et al. (2015)

A continuación presentamos un ejemplo de los criterios de adjudicación comúnmente utilizados para otorgar los procesos de licitación de evaluaciones de la Comisión Europea, en los cuales el 40 por ciento de la puntuación técnica se asigna a la composición del equipo.

Sin un desglose adicional del significado específico de cada criterio, la selección sigue siendo en cierto modo subjetiva. El ejemplo que mostramos a continuación (basado a grandes rasgos en una evaluación de Solidaridad Suiza en Haití, ver Groupe URD, 2013) se basa en un enfoque más estructurado.

Aspecto	Criterios	Puntuación
<b>Jefe del equipo</b>	Considerable experiencia en evaluación humanitaria y en el uso de distintas metodologías	15
	Dominio del francés o del criollo haitiano	5
	Excelentes habilidades de redacción en francés o inglés	10
	Amplia experiencia en Haití añade valor	5
<b>Todo el equipo</b>	Experiencia en la evaluación de intervenciones de reconstrucción o desarrollo	5
	Amplia experiencia en Haití, entre otras cosas una profunda comprensión de las estructuras económicas, el aparato institucional y el tejido social del país	10
	Gran fluidez en francés	10
	Capacidad de análisis de la información de encuestas cuantitativas (este requisito puede ser para un miembro del equipo que no viaje)	5
	Experiencia en intervenciones de desarrollo de capacidades o en la evaluación de programas de este tipo	5
	Experiencia en programas de apoyo a medios de vida o en la evaluación de intervenciones de medios de vida	5
	Experiencia en programas de reducción del riesgo de desastres (RRD)	5
	Considerable fluidez en criollo haitiano añade valor	5
	<b>Enfoque</b>	Grado al que el enfoque propuesto refleja una adecuada comprensión de los riesgos y restricciones
<b>Licitador</b>	Garantía de estabilidad en la composición del equipo a lo largo de todo el período de evaluación	10
	Experiencia en evaluaciones similares	10
<b>Precio</b>	Puntuación = 15x (coste de la oferta que califica más económica)/(coste de esta oferta)	15
<b>Máxima puntuación posible</b>		130

Se trata de un conjunto de criterios mucho más detallado, que facilita la calificación de las ofertas. La mayor parte de la puntuación técnica (85/115, o 74 %) se adjudica a las capacidades del equipo. Aunque posiblemente no deseen entrar en este nivel de detalle, es buena idea especificar qué capacidades se requieren tanto de la persona que dirige el equipo como del equipo. Este ejemplo podría ser mejorado aún más ofreciendo ejemplos de puntuaciones altas y bajas como los siguientes:

- Capacidad de saludar en criollo, 1 punto, capacidad de entrevistar en criollo sin asistencia o malentendidos, 5 puntos.
- Experiencia en Haití, 1 punto por año de experiencia.
- Dominio del francés: puede oscilar entre conocimiento básico, 1 punto, y lengua materna, 10 puntos.

Si un equipo consta de varias personas, estas pueden ser calificadas de manera individual, para luego promediar la puntuación, salvo que sea suficiente que un solo miembro del equipo posea la habilidad requerida.

Aspecto	Criterios	Puntuación total	Puntuación mínima para calificar
<b>Jefe del equipo</b>	Experiencia y capacidades de la persona responsable del equipo desagregadas en elementos principales, por ejemplo: experiencia en liderar una evaluación, experiencia geográfica, experiencia sectorial o institucional, etc. Para mayor transparencia, puede ser útil incluir una rúbrica de puntuaciones para los elementos individuales (por ejemplo, experiencia en liderar evaluaciones humanitarias – 1 punto por cada dos años hasta un máximo de 7 puntos, o dominio de idiomas (por ejemplo, lengua materna 4, fluidez 3, fluidez limitada 2, conocimiento básico 1).	20	15
<b>Entrevista al jefe del equipo</b>	Entrevista para establecer el liderazgo y las habilidades interpersonales de la persona responsable del equipo.	20	15
<b>Otros miembros</b>	Experiencia y capacidades de los demás miembros del equipo, desglosadas en principales áreas de experiencia y capacidad.	30	20
<b>Contexto</b>	Comprensión probada del contexto y de los problemas más importantes que entraña la evaluación.	10	5
<b>Enfoque</b>	Enfoque propuesto para el abordaje de la tarea de evaluación (podría incluirse más bien en la entrevista a la persona responsable del equipo).	20	10
<b>Puntuación técnica total/puntuación mínima para calificar</b>		100	65



### Consejo útil

No todos los criterios tienen la misma importancia. Consideren cuáles de sus criterios de evaluación resultan esenciales para garantizar una buena evaluación.

Se pueden aplicar criterios esenciales aplicando una puntuación mínima para cada criterio relevante. Por ejemplo, se puede decidir tomar en cuenta solo a los candidatos a jefes de equipo que hayan obtenido una puntuación mínima de 10 de 15 puntos para el criterio de experiencia previa en evaluación. Si se realiza una entrevista como parte del proceso de selección, una puntuación alta en la entrevista (reflejando las habilidades interpersonales de la persona candidata) puede ser un criterio esencial. Referencias favorables de clientes previos también pueden ser un criterio básico.

La puntuación para el contexto y el enfoque podría dividirse entre la expresión de interés y la entrevista con la persona candidata a jefe/a del equipo, cuando la decisión debe tomarse sobre esta base y no sobre la de una licitación formal.

Ver la subsección a continuación sobre [procedimientos de contratación](#) para una discusión del impacto del coste y por qué la oferta más económica no debe ser automáticamente seleccionada.<sup>1</sup>



### Consejo útil

Conseguir al consultor o consultores apropiados para una evaluación es importante y toma tiempo. Cuanto más temprano convoquen, mayores serán sus opciones.

El tipo de consultor varía, dependiendo de la tarea de evaluación. Para lidiar con temas delicados relacionados con alianzas institucionales, los consultores deberán ser competentes en relaciones institucionales en general, y asociaciones o alianzas en particular. Una ETR requiere consultores con experiencia operativa previa para tener credibilidad frente al personal de campo. Una evaluación sectorial específica puede requerir evaluadores con experiencia de trabajo en ese sector. Estos requisitos deben reflejarse en los criterios de selección del equipo de evaluación.

Entre las consideraciones a tomar en cuenta al seleccionar a los consultores se encuentran:

- **Qué tan bien conoce la organización el consultor.** Un consultor que ya conoce la organización necesitará menos tiempo para familiarizarse con ella, pero también puede inclinarse menos a cuestionar sus creencias predominantes.
- **Qué tan bien conoce el contexto el consultor.** Este aspecto es especialmente importante en situaciones de crisis y conflictos complejos. Dependiendo de la evaluación, puede privilegiarse el conocimiento ya sea del contexto o del tipo de actividad a evaluar.

- **Si el consultor posee las capacidades y experiencias necesarias en los diseños y metodologías de evaluación propuestos.** Esto puede incluir conocimiento de áreas sectoriales (por ejemplo, APS) y de la teoría apropiada (por ejemplo, teoría organizacional e ingeniería de procesos para evaluaciones de procesos institucionales).
- **La composición general del equipo.** Los equipos más numerosos pueden tener mayores capacidades que ofrecer pero también pueden ser más caros, necesitar mayor soporte logístico sobre el terreno, comportar un mayor riesgo de desavenencias internas y requerir un tiempo más largo para terminar el informe.

Otro factor a considerar es si reclutar a los consultores ya sea individualmente o en equipo a través de una empresa consultora. En el Cuadro 9.3 se resumen los pros y los contras de cada elección. En muchos casos, la decisión dependerá de las políticas institucionales o de las implicaciones en términos de oportunidad.<sup>2</sup>

**Cuadro 9.3:** Ventajas y desventajas de contratar individuos o empresas consultoras

Individuos		Empresa consultora	
+	Ustedes seleccionan a las personas y las capacidades que desean.	+	Hay un solo contrato que gestionar.
+	Pueden cerciorarse de que todos los miembros del equipo tengan algo que ofrecer.	+	Ustedes tienen mayor injerencia en los aspectos de calidad.
+	Puede ser menos complicado que un contrato.	+	Resolver las controversias dentro del equipo es responsabilidad de la empresa consultora.
+	Mantienen una relación directa con todos los integrantes del equipo.	+	Solo tienen que lidiar con unos cuantos pagos.
+	Esta opción suele ser menos costosa.	-	La empresa consultora asume algunos de los riesgos.
-	Limitaciones financieras pueden impedirles contratar a los consultores más capacitados.	-	Esta opción suele ser más cara.
-	Resolver las controversias dentro del equipo es vuestra responsabilidad.	-	El equipo puede tener un solo especialista experimentado, respaldado por personal con menos experiencia.
-	Tienen que dar soporte de recursos humanos al equipo.	-	Es posible que algunos miembros del equipo nunca antes hayan trabajado juntos.
-	Tienen que organizar la logística sobre el terreno.	-	Aún pueden tener que lidiar con la logística sobre el terreno y otros asuntos si la empresa consultora es incompetente.
-	Tienen que lidiar con muchos pagos diferentes.	-	Limitaciones financieras pueden impedirles contratar a un equipo de manera oportuna.
-	Ustedes asumen todos los riesgos.		

Fuente: Recopilación propia de los autores

En ocasiones los consultores pueden seleccionarse sobre la base de experiencias previas, pero darse ese lujo no siempre es posible. Se puede pedir a los candidatos que presenten notas breves sobre el enfoque que adoptarían, pero es preferible seleccionar a los consultores con base en su trabajo previo. ¿Quién escribió los informes de evaluación que consideran de buena calidad?



#### Consejo útil

No se guíen por la autoría de un candidato de una evaluación previa solamente – contacten al gestor de la evaluación y pregúntenle por el historial de trabajo del consultor. El informe final puede haber sido la versión 19 y haber sido reescrito de principio a fin por otro miembro del equipo.

La relación contractual dependerá de la política de contrataciones de la organización. Una práctica habitual consiste en condicionar los pagos a las diferentes etapas de la evaluación, por ejemplo, la firma del contrato, la aprobación del informe inicial, la culminación del trabajo de campo y la aprobación de los diferentes borradores y del informe final. Davidson (2012a) precisa que los términos del contrato deben permitir que el gestor de la evaluación rescinda el contrato si el equipo de evaluación demuestra ser incapaz de cumplir con sus obligaciones.



#### Consejo útil

Incluyan explícitamente en los términos del contrato el derecho a poner fin a la evaluación en la fase inicial si el informe inicial no fuera satisfactorio.

### Procedimientos de contratación

Si se invita a los evaluadores a presentar sus ofertas, el procedimiento habitual consiste en pedirles que presenten la oferta técnica y la oferta financiera por separado, cada una en un sobre sellado. Las ofertas técnicas se abren primero, y cada oferta recibe un puntaje por este rubro (ver la subsección [Selección del equipo externo adecuado](#) más arriba para un ejemplo de los criterios que podrían ser especificados). Los candidatos pueden recibir puntos después por la entrevista. Una vez calificadas las ofertas técnicas, se abren los sobres con las ofertas financieras y se hace el cálculo de la puntuación total.

Si la evaluación se otorga al equipo ganador sobre la base de una expresión de interés y una entrevista con el jefe del equipo, se puede invitar a los candidatos a presentar una oferta financiera en un sobre sellado, el mismo que se abrirá después de la entrevista. Este enfoque es especialmente indicado para evaluaciones pequeñas.

Existen tres modalidades principales para el otorgamiento de la evaluación:

- De las ofertas que califican técnicamente, se selecciona la de menor precio. Las ofertas que obtuvieran un nivel mínimo definido de puntos a nivel técnico y que no hubieran sido descalificadas de acuerdo a alguno de los criterios clave (por ejemplo, no obtener ninguna puntuación para un criterio esencial, como experiencia en evaluación) se considerarán técnicamente calificadas. Este es un buen enfoque para el caso de adquisiciones o servicios simples, pero no para seleccionar un servicio que exige un alto nivel de competencias técnicas complejas, como lo es la evaluación.
- Sobre la base de la puntuación técnica más una puntuación económica para diferenciar entre la oferta A y la oferta más baja que califica. Este es quizás el mejor enfoque ya que pone el mayor énfasis en las capacidades del equipo de evaluación.
- Sobre la base de la puntuación técnica, multiplicado por una puntuación económica para la diferencia entre la oferta A y la oferta más baja que califica.

Si se multiplica la puntuación técnica total por el precio, la selección dependerá básicamente del precio. En este caso, aplicar la primera opción hace que la oferta ganadora sea la Oferta 2, pese a haber obtenido solo un 73 % en el puntaje técnico, frente al 90 % de la Oferta 1.

Restringir el precio a solo algunos de los puntos es un enfoque más adecuado que es utilizado por UNICEF y los donantes escandinavos.

Elemento		Oferta 1	Oferta 2
Porcentaje de puntuación técnica máxima	A	90 %	73 %
Precio relativo (en comparación con el más bajo)	B	1.25	1
Factor precio (1 / B)	C	0.8	1
Fórmula de la UE (A x C)		72 %	73 %
Puntuación técnica (80 % entre A)	D	72 %	58 %
Puntuación económica (20 % x C)	E	16 %	20 %
Fórmula de UNICEF (D + E)		88 %	78 %

## 9.6 Desafíos para el liderazgo y el trabajo en equipo

Como cualquier otro proyecto, una EAH tiene que ser correctamente gestionada para tener éxito. Es por eso que Grieg (2011: 3) incluye sólidas capacidades de gestión de proyectos como una competencia clave para los evaluadores, y esto incluye:

1. Competencia técnica;
2. Fuertes capacidades de gestión de proyectos;
3. Excelente capacidad de análisis;
4. Excelente capacidad de comunicación con las partes interesadas;
5. Flexibilidad;
6. Orientación a la colaboración.

Las últimas tres son habilidades interpersonales. En la EAH, la competencia técnica incluye conocimiento de los principios humanitarios y del sistema humanitario, así como capacidad de trabajo bajo presión o en entornos inseguros.

Existen tres potenciales niveles de gestión: gestión a través del departamento de evaluación de la organización que encarga la evaluación (el gestor de la evaluación), gestión por el administrador del contrato cuando se emplea una empresa consultora, y gestión del proceso y del equipo por el jefe del equipo. Cada uno de estos niveles se ocupa de diferentes temas. La presente subsección se centra en la gestión del proceso de EAH y del equipo de evaluación por parte del jefe del equipo.

Incluso cuando los evaluadores son competentes, se tienen los recursos adecuados y los objetivos están claramente definidos, una buena planificación sigue siendo esencial para el éxito de una evaluación. Prácticamente todas las EAH están sujetas a restricciones de tiempo. Aprobar los TdR, reclutar al equipo, redactar y circular el informe en borrador e incorporar los comentarios de los revisores casi siempre toma más tiempo de lo esperado; el trabajo de campo con frecuencia es el que más sufre cuando se producen retrasos o la evaluación excede el presupuesto. Una planificación cuidadosa puede ayudar a evitar estos contratiempos.



### Consejo útil

Aclaren desde el principio qué apoyo dará la organización que encarga la evaluación al equipo de evaluación. ¿El equipo tendrá acceso a los archivos, espacio de trabajo en la oficina central y en las oficinas de campo, apoyo de la oficina central y las oficinas de campo con la concertación de las citas, servicios de seguridad, transporte y asistencia con las reservas de alojamiento en terreno?

La magnitud de la evaluación determina el tamaño del equipo. Como se señaló anteriormente, sin embargo, los equipos de mayor tamaño comportan problemas y riesgos que pueden superar los beneficios de contar con una gama de competencias más amplia. Los equipos más numerosos suponen más trabajo para el jefe del equipo, por ejemplo, con la coordinación de las tareas y el cotejo de los insumos de los diferentes miembros del equipo. Asimismo, representan un problema en entornos inseguros. En algunos casos es difícil conseguir alojamiento satisfactorio en ciertos lugares y los equipos se tienen que dividir, dificultando las conversaciones informales fuera de horas de trabajo.



#### Consejo útil

Elaboren plantillas de documentos como listados de personas entrevistadas, bibliografías e itinerarios, formateados en el estilo en que desean recibir la información. Recibir los insumos en un formato estandarizado reduce el trabajo de convertirlos y cotejarlos.

Los equipos de EAH suelen ser convocados ya sea por una empresa consultora o por contratación directa. En cualquier caso, es posible que el jefe del equipo no haya trabajado antes con todos los miembros del equipo.



#### Consejo útil

Si están trabajando con algunos evaluadores por primera vez, y no los conocen por reputación, organicen el trabajo de campo de manera tal que puedan pasar un día juntos al principio de la evaluación, para formarse una idea de sus fortalezas y debilidades, así como de posibles sesgos.

Incluso los mejores evaluadores tienen “días malos”, y un buen desempeño previo no garantiza un buen desempeño en relación con la evaluación actual. Un miembro del equipo puede estar atravesando problemas personales, por ejemplo la enfermedad de un familiar cercano. Los miembros del equipo que están teniendo dificultades personales pueden necesitar el apoyo del responsable del equipo. Sin embargo, éste debe establecer plazos para las tareas y hacer seguimiento de inmediato en caso de incumplimiento. Esto puede ayudar a identificar y subsanar cualquier problema de desempeño con celeridad.

En ocasiones surgen problemas por conflictos de personalidad, problemas de desempeño o valores divergentes. Algunos entornos de una EAH pueden ser agobiantes. De estos problemas, los que se refieren al desempeño son los más fáciles de manejar – aunque para cuando salen a relucir puede ser muy tarde para prescindir de la persona o reemplazarla sin abandonar la evaluación.



### Consejo útil

Si un evaluador deficiente no puede ser retirado del equipo, emparejenlo con un mejor miembro del equipo para reducir en la mayor medida posible los problemas de calidad.

Los equipos grandes tienen mayores probabilidades de experimentar controversias en torno a los hallazgos, ya sea debido a diferencias de perspectiva o de percepción del contexto o entre las áreas en las que los miembros del equipo se han enfocado. Las siguientes opciones constituyen estrategias probadas y comprobadas para manejar equipos de evaluación numerosos:

- Dispongan un espacio al comienzo del trabajo de campo para que el equipo se conozca y para dilucidar los roles y expectativas. Consideren formas en que cada miembro del equipo pueda dar a conocer su área de especialización. Incluso cuando el tiempo escasea, invertir en cultura de equipo y en comunicación entre los miembros del equipo desde un inicio puede rendir frutos.
- Cerciórense de que todos los miembros del equipo vivan en un mismo lugar, para que se enfrasquen en conversaciones informales y socialicen entre sí.
- Utilicen un cuadro de evidencia (ver la [Sección 16, Análisis](#)) para llevar un registro de la evidencia y de los hallazgos emergentes, y compártanlo para mantener informados a los miembros del equipo y desarrollar una visión común.
- Si en un país hay un equipo numeroso, incorporen al programa de trabajo un espacio para la reflexión, de modo tal que el equipo pueda discutir los hallazgos que vayan emergiendo.
- Planifiquen reuniones del equipo después del trabajo de campo y antes de finalizar el primer borrador del informe, para discutir la evidencia, los hallazgos y las conclusiones y recomendaciones.

En ocasiones un miembro del equipo puede estar particularmente interesado en un tema en especial que, aunque relevante, es de importancia periférica en relación con el foco principal de la evaluación, o es muy estrecho o muy complejo como para ser incluido detalladamente en el informe principal. En dichos casos, una solución podría ser pedir al miembro del equipo que escriba sobre el tema en un anexo.

## 9.7 Gestión de controversias

Las disparidades entre la escala de la tarea de evaluación y los recursos disponibles también pueden generar controversias entre el gestor de la evaluación y el equipo de evaluación. Muchos de los desacuerdos de este tipo se originan por diferencias de percepción. Un informe inicial puede reducir el riesgo de malentendidos en una etapa temprana, pero las controversias también pueden surgir en la fase de elaboración del informe, llevando a múltiples revisiones y aumentando la frustración en ambos lados.



### Consejo útil

Una forma de reducir las desavenencias en una evaluación controvertida es formando un comité de coordinación que preste asesoría con relación a la aceptabilidad de los primeros borradores del informe (aunque la decisión final descansa sobre el gestor de la evaluación). Los comités de coordinación pueden contribuir asimismo a garantizar la calidad de todo el proceso de evaluación.

Los conflictos de personalidad son otra potencial fuente de controversia. Un medio para resolverlos puede ser contar con una política formal de solución de controversias, que especifique qué medidas se tomarían en caso de controversia e identifique quién actuaría de árbitro de ser necesario.



### Ejemplo de buena práctica: Creación de una política de solución de controversias

La Coalición de Evaluación del Tsunami adoptó una política de solución de controversias en caso de que éstas surgieran entre los miembros de los equipos de evaluación, entre un equipo y el comité de coordinación, entre los miembros del equipo de síntesis y entre el equipo de síntesis y el comité de gestión medular. Dicha política establecía que el autor principal era el responsable de manejar las relaciones entre los miembros del equipo. Si surgía una controversia seria entre los miembros del equipo, o entre un miembro del equipo y el jefe del equipo, la misma que el autor principal no pudiera resolver, el comité de gestión medular debía pedir al director de la Secretaría de ALNAP que preparase un informe dando cuenta de ambos lados de la disputa. El director de la Secretaría de ALNAP podría elaborar el informe por su cuenta o contratar a un evaluador experimentado para hacerlo. El comité de gestión podría entonces tomar la decisión de solicitar al autor principal que hiciera una de las siguientes cosas:

- Presentar solo una de las interpretaciones;
- Incluir ambas interpretaciones en el informe;
- Incluir solo una interpretación, pero dejar constancia de que la misma no era unánime.

Si los miembros del equipo de síntesis no estuvieran satisfechos con la resolución del problema, tendrían derecho a pedir el retiro de su nombre del informe.

Fuente: UNICEF (2010)



### Consejo útil

Las políticas formales de solución de controversias son apropiadas no solo para evaluaciones a gran escala sino también para cualquier evaluación que las pudiera necesitar.

# 10 / Métodos de revisión documental

Esta sección se centra en el uso de métodos de revisión documental en una evaluación y analiza cuánto tiempo debería dedicarse a la revisión documental inicial. Muestra, además, varias formas de presentar la información obtenida de una revisión o un estudio documental, de manera concisa, tanto para el equipo de evaluación como para las sesiones informativas y los informes de evaluación.

Las revisiones y estudios documentales recurren a métodos documentales para resumir la documentación.

## Las preguntas de evaluación determinan la revisión documental

Como sucede con las metodologías de campo, la amplitud y la profundidad de una revisión, así como la selección de los métodos de revisión documental, dependen de las preguntas de evaluación. Algunas preguntas de evaluación solo pueden ser respondidas mediante revisión documental, mientras que otras lo pueden ser de manera óptima por ese medio. Por ejemplo, una pregunta de evaluación que indaga si una intervención está alcanzando sus metas podría ser óptimamente contestada revisando los informes de seguimiento y luego triangulando la información con entrevistas de campo.

Ver la [Sección 6](#) para mayores detalles sobre las preguntas de evaluación.

## 10.1 ¿Qué es una revisión documental?

Una revisión documental es una revisión de uno o más documentos. Puede llevarse a cabo:

- Como parte de la [fase inicial](#), para esclarecer la tarea de evaluación o responder a preguntas específicas;
- Como parte del ejercicio de definición de los alcances de la evaluación para la elaboración de los [TdR](#);
- Durante el [trabajo de campo](#).

En algunos casos, la revisión documental constituye la evaluación en sí, como sucedió con la revisión del desempeño del fondo de respuesta a emergencias del Consorcio de Agencias Humanitarias Británicas (Stoddard, 2011) o la revisión documental de las consecuencias fortuitas de la asistencia noruega (Wiig y Holm-Hansen, 2014).<sup>3</sup>

Las revisiones documentales suelen servir también de base para el desarrollo de estudios sobre las enseñanzas aprendidas, como los estudios de ALNAP sobre las enseñanzas aprendidas de las respuestas a terremotos (Cosgrave, 2008), desastres urbanos (Sanderson et al., 2012) e inundaciones (Cosgrave, 2014).

Dichos estudios pueden basarse en evaluaciones, como en el caso de la revisión del Banco Mundial de evaluaciones pasadas como insumo para mejorar la respuesta a las inundaciones en África Occidental (IEG, 2010), o bien en grupos de documentos más amplios, como la revisión del Centro de Estudios sobre Refugiados de las enseñanzas aprendidas tras una década de discusiones sobre los refugiados en situación de desplazamiento prolongado (Milner y Loescher, 2011).

Las revisiones documentales pueden ser estructuradas o no estructuradas. Las primeras emplean una estructura formal para el análisis documental, mientras que las segundas consisten básicamente en la lectura de antecedentes.



### Consejo útil

La información, porque esté escrita no significa que sea automáticamente exacta o confiable. Este tipo de información en ocasiones es considerada como más exacta o más confiable que la información obtenida en entrevistas. Pero ese no es necesariamente el caso, y toda la información proveniente de documentos o entrevistas debe ser validada mediante triangulación.

Una revisión documental comprende lo siguiente:

- Identificación de los documentos. Normalmente, la organización que encarga la evaluación identifica un grupo inicial de documentos, pero dependerá del equipo de evaluación identificar materiales adicionales.
- Lectura o incluso análisis de la documentación.

La revisión documental puede ser realizada por:

- El equipo de evaluación, ya sea como una tarea inicial menor o como una parte importante de la evaluación. La revisión de la documentación para la segunda fase de la evaluación del proyecto de vinculación de ayuda de emergencia, rehabilitación y desarrollo (LLRD) tras el terremoto y los tsunamis del Océano Índico, es un ejemplo de este tipo de revisión (Cosgrave et al., 2009).
- Consultores especialmente contratados por la organización que encarga la evaluación. Esto se hizo para la evaluación del proyecto de LLRD en la evaluación de la respuesta al terremoto y a los tsunamis de 2004 en el Océano Índico (Buchanan-Smith y Fabbri, 2005). En este ejemplo, los consultores no solo identificaron la documentación sino que también la analizaron.
- La organización que encarga la evaluación, aunque esto no es lo usual, salvo que la revisión forme parte de un ejercicio para definir los alcances de la evaluación. La participación de la organización que encarga la evaluación suele terminar con la identificación de los documentos más importantes que deben ser sometidos a consideración del equipo de evaluación.

## 10.2 ¿Por qué hacer una revisión documental?

Las revisiones documentales representan una opción económica para que la evaluación:

- Capitalice el conocimiento adquirido a raíz de evaluaciones previas y otras investigaciones. Por ejemplo, la evaluación de Solidaridad Suiza de su respuesta a los tsunamis de 2004 en el Océano Índico se basó, para el cuestionario de su encuesta preliminar, en un análisis de estudios de impacto previos (Ferf y Fabbri, 2014: 17).
- Capitalice el conocimiento recogido en los documentos de seguimiento del proyecto.
- Permita al equipo de evaluación entender el contexto rápidamente.
- Identifique temas potencialmente decisivos para el trabajo de campo posterior. La revisión temática del Programa Mundial de Alimentos (PMA) de su programa de alimentación materno-infantil se sustentó en una revisión documental realizada en 2002, la misma que identificó diferencias clave en las operaciones del PMA en diferentes regiones (Hoogendoorn et al., 2005: 10-11).
- Identifique potenciales criterios para los juicios de valor, las fuentes y las metodologías de la matriz de evaluación.

Por las razones arriba expuestas, una revisión documental debería formar parte integrante de toda evaluación. Incluso una evaluación de bajo presupuesto necesitará que el equipo al menos revise los documentos del proyecto y cualquier otra evaluación previa pertinente.

Las revisiones documentales son particularmente útiles en la EAH para:

- Establecer una cronología de qué ocurrió y cuándo. Son indispensables si la evaluación ha de tomar en cuenta los aspectos como oportunidad y/o eficacia.
- Demostrar cómo han cambiado las prioridades de la intervención a través del tiempo. Las crisis humanitarias son inestables y el contexto puede cambiar rápidamente. Debido a la alta rotación de personal clave que caracteriza a las crisis humanitarias, el equipo de gestión actualmente en funciones puede no saber por qué se han adoptado ciertos enfoques. Una revisión documental puede ayudar al equipo de evaluación a entender mejor la evolución de la respuesta.
- Entender diferentes puntos de vista. Este es el caso especialmente en emergencias complejas, en que el personal humanitario puede identificarse con la población a la que están asistiendo. Una revisión documental puede ayudar al equipo de evaluación a entender de una manera más global el contexto.
- Identificar las enseñanzas aprendidas a raíz de operaciones previas.

## 10.3 ¿Cuánto tiempo debe durar una revisión documental?

El tiempo requerido para la revisión documental dependerá de:

- La pregunta o preguntas de evaluación y si los documentos son fuentes de información útil. Si se espera que los documentos sean un recurso clave para responder a las preguntas, en ese caso se deberá asignar más tiempo a la revisión.
- La naturaleza y los alcances de la evaluación. Si la evaluación está examinando un período de intervención extenso, los documentos tenderán a ser importantes fuentes de información y se requerirá más tiempo para la revisión.
- La cantidad y calidad de la documentación disponible. Las crisis recientes pueden contar con escasa documentación, más allá de informes de situación y diagnósticos de necesidades. Las crisis de larga duración pueden generar miles de documentos. Si el conjunto de documentos es extenso, se deberá asignar más tiempo a la revisión.
- La disponibilidad de resúmenes de análisis publicados. Si estos ya existen, es posible que se requiera menos tiempo para la revisión documental.

## 10.4 Llevando a cabo una revisión documental

El equipo de investigación en evaluación de los Centros para el Control y la Prevención de Enfermedades (CDC) ofrece consejos sucintos sobre revisiones documentales para una evaluación (Equipo de Investigación en Evaluación, 2009). Hagen-Zanker y Mallett (2013) ofrecen orientación sobre revisiones literarias rigurosas en el sector de desarrollo internacional.

### Paso 1: Identifiquen posibles fuentes

Las fuentes documentales dependen de las preguntas de evaluación y del tiempo disponible para la revisión. Pueden incluir:

- Documentos clave mencionados en los TdR. Estos pueden comprender no solo documentos específicamente referidos al proyecto o programa sino también documentos estratégicos más generales;
- Fuentes específicas de la organización. Estas pueden comprender los documentos centrales vinculados a la evaluación (algunas organizaciones que encargan evaluaciones facilitan dichos documentos en un CD-ROM o los cuelgan en una carpeta en línea). Los portales web o las intranets institucionales pueden ser una rica fuente de diagnósticos de necesidades, informes situacionales, propuestas de proyectos e informes de seguimiento;
- Relief Web, portales web sobre crisis específicas (por ejemplo, el Portal de ALNAP sobre Haití)<sup>4</sup> y portales temáticos (por ejemplo, el Portal sobre Respuestas Humanitarias Urbanas);<sup>5</sup>



- El Portal de ALNAP sobre Evaluación y Aprendizaje Humanitarios (HELP);<sup>6</sup>
- Referencias y bibliografías existentes;
- Búsquedas web;
- Búsquedas avanzadas de Google Académico;
- Búsquedas en bases de datos académicas.

#### **Consejo útil**

Elaboren un listado de las características de los documentos que están buscando. Pueden decidir centrarse inicialmente en evaluaciones y revisiones, o después de una fecha en particular, en materiales elaborados – hacerlo puede contribuir a acelerar el proceso de revisión.

## **Paso 2: Categoricen sus documentos**

No todos los documentos merecen el mismo peso o la misma atención. Puede ser útil dividir los documentos disponibles en niveles o categorías:

### **Nivel 1**

Documentos centrales que enmarcan el sujeto de evaluación y a menudo se encuentran mencionados en los TdR. Entre estos se encuentran los principales documentos estratégicos y de planificación así como evaluaciones previas del programa. Normalmente hay entre cinco y 20 documentos centrales, aunque las evaluaciones de gran alcance, que abarcan varios países y sectores, así como largos periodos de tiempo, pueden tener más.

### **Nivel 2**

Documentos específicamente referidos al sujeto de la evaluación, tales como informes de situación, informes de avance, informes de seguimiento y propuestas de proyectos (cuando la evaluación es de un programa más amplio). Su número puede oscilar entre 20 y varios miles de documentos, dependiendo del alcance de la evaluación.

### **Nivel 3**

Documentos de antecedentes, entre ellos artículos de prensa e informes de otras organizaciones. Las publicaciones de ReliefWeb constituyen un ejemplo de este tipo de documento, pudiéndose colgar más de 700 publicaciones por semana sobre una nueva crisis, lo cual significa que puede haber desde menos de cien hasta decenas de miles de documentos, dependiendo del alcance de la evaluación.

### Paso 3: Decidan si necesitan un enfoque estructurado

Muchas evaluaciones se circunscriben al primer nivel de documentación y un enfoque no estructurado resulta por lo tanto apropiado. Una revisión documental que incluye documentos del segundo y tercer nivel requiere un enfoque estructurado, por ejemplo:

- Una herramienta de observación estructurada para registrar los comentarios (ver ejemplo más adelante);
- Uso de rúbricas para asignar una puntuación a ciertos aspectos de la documentación (ver ejemplo más adelante);
- Indexación y búsqueda de documentos para el análisis contextual.

### Paso 4: Armonicen la estrategia de revisión documental con el nivel requerido

El equipo de evaluación debe leer los documentos primarios lo antes posible. Las estrategias para el siguiente nivel pueden incluir el uso de una rúbrica y una herramienta para anotar las observaciones del revisor. Los documentos del tercer nivel pueden estar sujetos a un análisis de contenido.



#### Consejo útil

Utilicen un *software* bibliográfico, ya sea productos comerciales como Endnote, productos inicialmente gratuitos como Mendeley o productos gratuitos como Zotero, lo cual es imprescindible si deben examinar muchos documentos. Estos programas permitirán que las referencias de las bibliotecas digitales sean compartidas entre los miembros del equipo y garantizan que los documentos sean citados de manera homogénea.



#### Consejo útil

Compartan los documentos en una nube a través de una aplicación como Dropbox, Box, One Drive o Google Drive, a fin de almacenar una copia de todos los documentos y ponerla a disposición de todos los miembros del equipo.

## 10.5 Herramientas para revisiones documentales

Las herramientas básicas para revisiones documentales incluyen formularios para registrar de manera estructurada los comentarios de la revisión. Estos pueden usarse en combinación con rúbricas para asignar una puntuación a los documentos en relación con aspectos específicos.

### Rúbricas

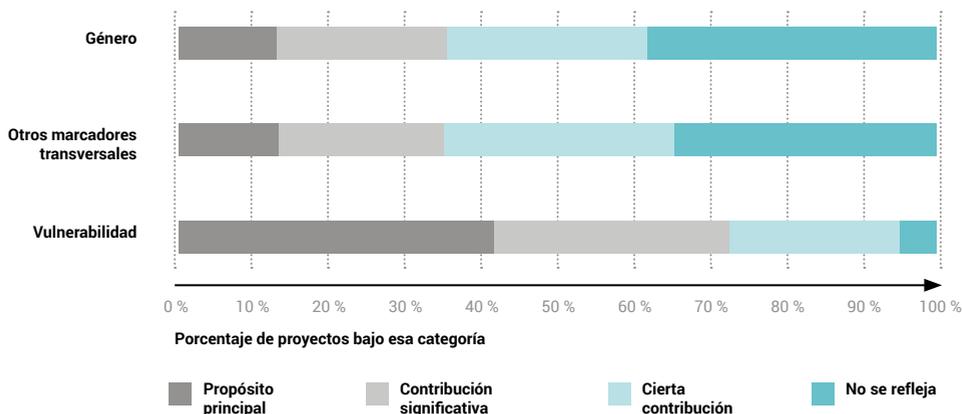


#### Definición: Rúbrica

Una rúbrica es una herramienta de calificación que se usa para evaluar un documento en función de una serie de criterios de manera consistente.

Las puntuaciones asignadas pueden emplearse para categorizar los documentos o comparar los cambios de énfasis a través del tiempo. Por ejemplo, en la evaluación quinquenal del Fondo Central de Respuesta a Emergencias (CERF) se emplearon rúbricas para las áreas de género, vulnerabilidad y otros temas transversales, arrojando que, mientras que el 95 por ciento de los proyectos planeaban realizar al menos alguna contribución a la reducción de la vulnerabilidad, menos de dos tercios planeaban realizar alguna contribución a la igualdad de género u otros temas transversales (Channel Research, 2011).

Figura 10.1: Género, vulnerabilidad y marcadores transversales



\*Marcadores de género, vulnerabilidad y otros temas transversales para una muestra aleatoria de 258 proyectos financiados por el CERF en 16 países del estudio; 41 proyectos fueron excluidos porque no contaban con información suficiente o suministraban servicios comunes.

El Marcador de Género (IASC, 2010) es un ejemplo de una rúbrica. Se utiliza principalmente para evaluar el diseño de un proyecto.

Nivel	Descripción
0	Ningún componente del proyecto incide en la igualdad de género.
1	El proyecto está diseñado para contribuir de una manera limitada a la igualdad de género. Las dimensiones de género están incluidas de manera significativa en solo uno o dos de los tres componentes básicos: diagnóstico de necesidades, actividades y efectos.
2a	<b>Transversalización del enfoque de género:</b> El proyecto está diseñado para contribuir significativamente a la igualdad de género. Las necesidades diferentes de las mujeres y las niñas frente a las de los hombres y los niños han sido analizadas e integradas adecuadamente en los tres componentes básicos: diagnóstico de necesidades, actividades y efectos.
2b	<b>Acciones focalizadas:</b> El propósito principal del proyecto es impulsar la igualdad de género. Todo el proyecto ya sea (a) se focaliza en las mujeres o los hombres, las niñas o los niños que tienen necesidades especiales o son objeto de discriminación, o (b) focaliza todas sus actividades en el desarrollo de servicios específicos de género o en desarrollar relaciones más equitativas entre las mujeres y los hombres.

Existe un creciente interés en el uso de rúbricas en la evaluación en general, y no solo en las revisiones documentales (Oakden, 2013a; 2013b; Rogers, 2013). Ver el uso de minirúbricas en talleres por Davidson (2014).

Se pueden desarrollar rúbricas similares para cualquier aspecto relevante. Una rúbrica puede utilizar una escala más compleja o limitarse a una escala simple de cuatro puntos para calificar el nivel de atención a cualquier tema, por ejemplo discapacidad, mitigación de la corrupción, RRD o derechos de la infancia. Una escala simple de cuatro puntos puede dividir los documentos en:

- Documentos que no inciden en el tema para nada.
- Documentos que inciden tangencialmente en el tema. La rúbrica debe indicar qué significa “tangencialmente”.
- Documentos que inciden de una manera significativa en el tema. La rúbrica debe especificar qué quiere decir “de alguna manera significativa”.
- Documentos cuya focalización principal es el tema.

Las rúbricas resultan útiles para garantizar lo siguiente:

- La coherencia entre distintos revisores;
- La coherencia a través del tiempo en el caso de un mismo revisor.

Es buena idea contar con una segunda persona para revisar una selección aleatoria de documentos ya calificados, con el fin de identificar posibles sesgos o calificaciones inconsistentes.

## Formularios de revisión estructurada

Se trata de formularios sencillos donde el revisor podrá anotar los aspectos centrales de un documento para hacerle seguimiento más adelante. Por ejemplo, si el interés recae sobre la coordinación tanto dentro de un sector como entre diferentes sectores, podríamos revisar los documentos con un análisis rápido inicial y anotar los aspectos de interés, tal como se muestra a continuación.

Otra forma de realizar una revisión estructurada es utilizar la herramienta de evidencia que se describe en la Sección 16, Análisis. Los formularios de revisión estructurada se pueden usar asimismo para registrar las puntuaciones asignadas en rúbricas a diferentes aspectos.

Doc.	Pág.	Coordinación sectorial	Coordinación intersectorial
1	6	“Asistencia regular a las reuniones del clúster”	
	9	“Implementamos de inmediato la decisión del clúster de WASH de clorar toda el agua de los camiones cisterna en la fuente y controlar los niveles de cloro en los puntos de descarga.”	
	19	“El 87 % de los camiones cisterna cumplieron con la norma del clúster. Aquellos que incumplieron la norma continuamente fueron retirados del contrato y se dio cuenta de sus números de placa a las otras organizaciones que forman parte del clúster.”	“El equipo de Agua Potable y Saneamiento se reunió con el equipo central del clúster de logística para coordinar la eliminación de residuos sólidos.”

## Análisis de contenido

El análisis de contenido es una herramienta útil para manejar un gran número de documentos. La mayoría de los principales paquetes para el análisis de información cualitativa incluyen medidas para realizar un análisis de contenido pero no siempre son fáciles de usar. Una alternativa más económica y sencilla es emplear una herramienta de indexación digital como dtSearch<sup>7</sup> para buscar documentos a partir de palabras clave específicas o llevar a cabo otros tipos de análisis de contenido.



### Definición: Análisis de contenido

El análisis de contenido es un análisis de información textual que se realiza de manera estandarizada, de modo tal que los evaluadores puedan hacer inferencias con respecto a la información.

El análisis de contenido tradicionalmente adopta la forma de documentos de codificación, reduciendo el texto a un conjunto de categorías. Una modalidad de análisis de contenido que ha sido empleada en varias EAH, entre ellas la evaluación quinquenal del CERF (Channel Research, 2011), es el análisis de palabras clave.



### Definición: Análisis de palabras clave

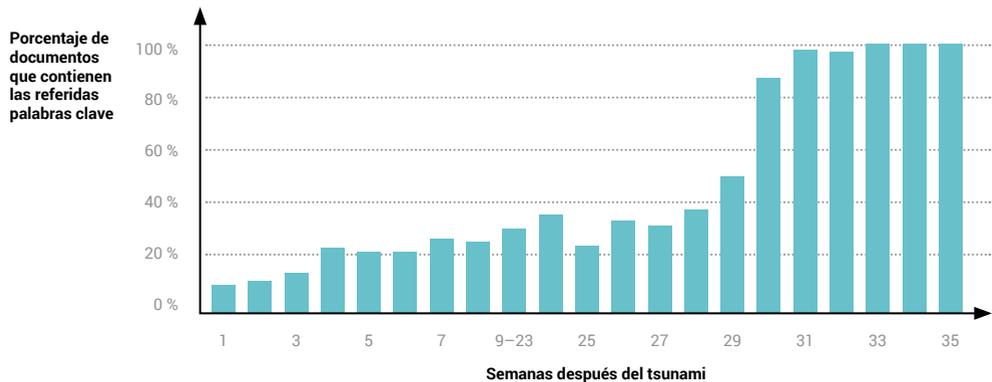
El análisis de palabras clave es una modalidad de análisis de contenido que examina con qué frecuencia ocurren ciertas categorías de efectos ya sea para sacar a la luz tendencias en el tiempo de un conjunto de documentos o comparar dos grupos de documentos.

Un listado de palabras clave puede tener hasta 500 términos. Algunas categorías se encontrarán representadas por varias palabras clave. Por ejemplo, entre las palabras clave para la categoría de agua potable y saneamiento (WASH, de su nombre en inglés) se encuentran las siguientes: pozo vertical, cloro, defecación, heces, higiene, letrina, excremento humano, saneamiento, jabón, manantial, inodoro, lavado, agua, abastecimiento de agua, agua y saneamiento, pozos.

Rara vez se utilizan otras modalidades de análisis de contenido en la EAH. Una excepción fue la evaluación de impacto de la asistencia para la integración de los refugiados que retornaban a su país de origen en Burundi, donde las entrevistas transcritas fueron sometidas a un análisis de contenido (Telyukov et al., 2009).

Tres guías provechosas para el análisis textual son: Krippendorff (2004), sobre textos estándar; Benini (2009), una breve guía para trabajadores humanitarios y de desarrollo sobre cómo usar tres herramientas de análisis textual; y la guía de la Oficina de Contabilidad General de los Estados Unidos (GAO, 1996). El análisis de contenido puede usarse para demostrar el surgimiento y la evolución de diferentes conceptos a través del tiempo. El ejemplo a continuación muestra la prevalencia de términos relacionados con capacitación en el conjunto de documentos de ReliefWeb para la evaluación de la Coalición de Evaluación del Tsunami del terremoto y los tsunamis de 2004 en el Océano Índico. El análisis demuestra que, al cabo de 30 semanas, prácticamente todos los documentos se referían a capacitación de alguna manera, pero que inicialmente se prestó escasa atención al tema.

**Figura 10.2:** Prevalencia de palabras clave relacionadas con capacitación en los 14,560 documentos sobre el tsunami



## 10.6 Resumiendo la información de la revisión documental

Los hallazgos de las revisiones documentales efectuadas por el equipo de evaluación pueden ser incluidos en las sesiones informativas, el informe inicial y el informe de evaluación principal. Un desafío particular es comprimir la gran cantidad de información que resulta de una revisión documental en un formato accesible. Esto puede hacerse en forma de cuadros y gráficos. Los resúmenes informativos creados de esta manera pueden servir también de insumos para enriquecer la información de todo el equipo de evaluación.

### Cronologías

Las cronologías son muy útiles en la EAH, ya que pueden ayudar a decidir la secuenciación de la asistencia.

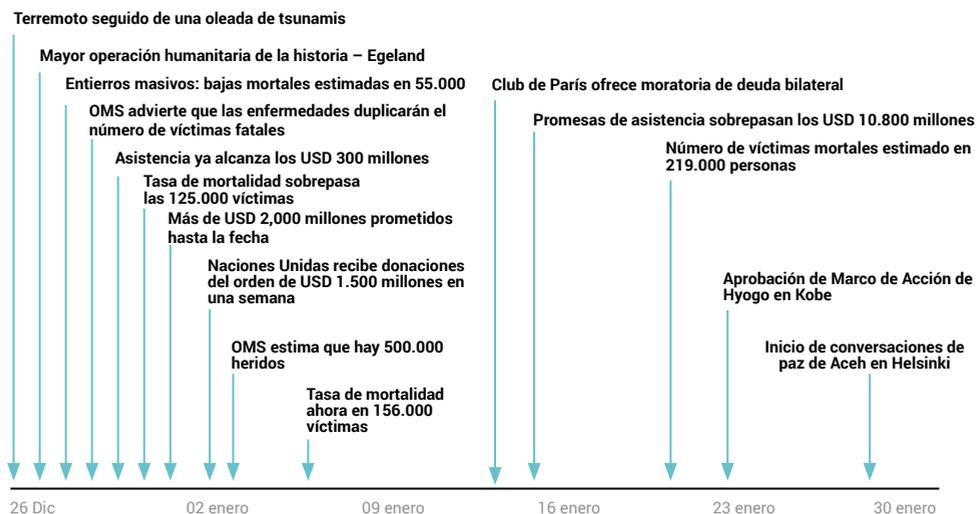


#### Consejo útil

Empecen a recopilar información para la cronología lo antes posible. Empezar temprano con la cronología puede ayudar a situar diferentes interpretaciones en contexto a medida que inician la revisión documental.

Las cronologías pueden presentarse en forma tabular o gráfica. La cronología parcial que presentamos a continuación ilustra el inicio de la respuesta a los tsunamis de 2004 en el Océano Índico y muestra que se comprometieron fondos mucho antes de que existiera un diagnóstico de necesidades detallado.

Figura 10.3: Cronología parcial del terremoto y los tsunamis en Asia el 26 de diciembre de 2004



Las cronologías pueden sustentarse en las fechas de la documentación analizada.



### Consejo útil

Fecha los documentos añadiendo un prefijo al nombre de cada archivo del conjunto de datos conforme al siguiente formato: AAAmmDD. Esto permite clasificar rápidamente los documentos y constatar qué documentos fueron publicados en qué fecha.

Un *software* de indexación o de análisis contextual, por ejemplo el programa dtSearch arriba citado, puede servir para buscar las fechas de los documentos, ahorrando tiempo para la elaboración de la cronología.



### Consejo útil

Involucren a todo el equipo en la elaboración de la cronología, ya que esto contribuirá a que todo el equipo entienda mejor la forma en que evolucionó la respuesta.

## Otros tipos de información cronológica

Las cronologías son solo una forma de información cronológica, que puede ayudar al equipo a entender cómo se desarrolló una crisis. Es posible, por ejemplo, contrastar los eventos con una curva epidémica, como se hizo con el financiamiento del CERF y la epidemia del dengue de 2009 en Cabo Verde.

**Figura 10.4:** Línea de tiempo de la epidemia del dengue en Cabo Verde y proceso de solicitud de un préstamo del CERF

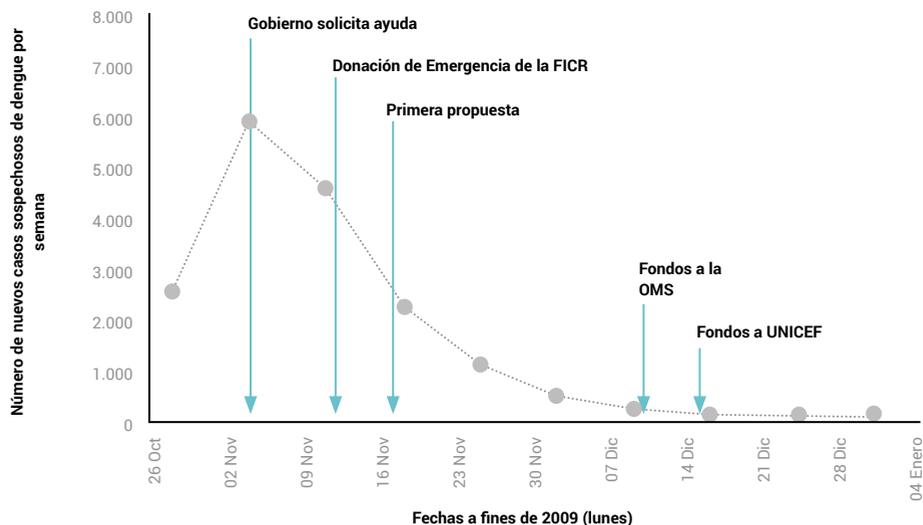
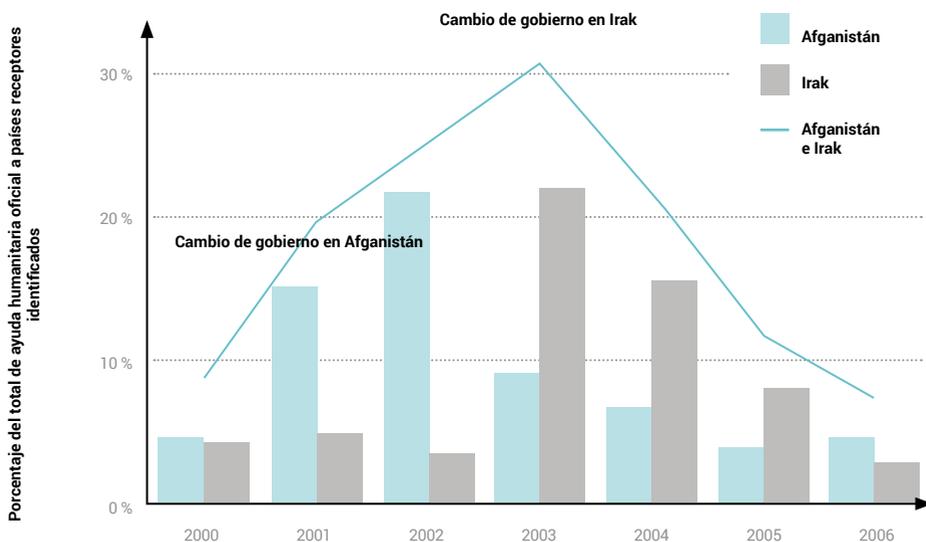


Figura 10.5: Asistencia humanitaria para Afganistán e Irak 2000-2006

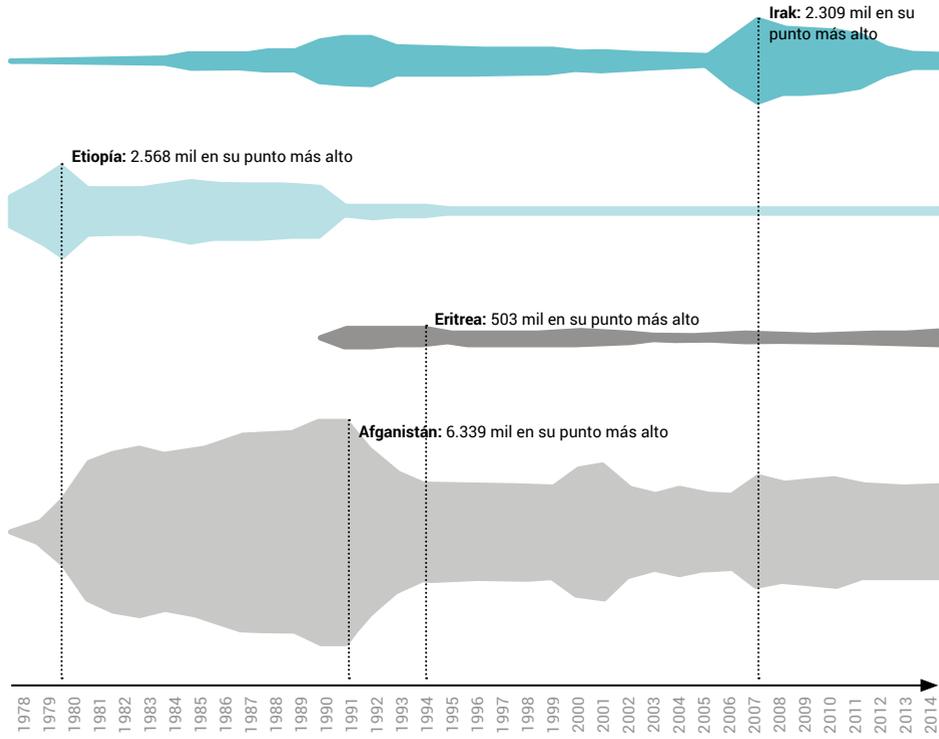


Fuente: Gráfico 2a, OCDE/CAD, 10 de abril de 2008. Dos terceras partes de la asistencia humanitaria se destinaron a países receptores identificados durante este período.

La información cronológica es útil asimismo para las evaluaciones de política. El gráfico a continuación revela cómo el cambio de gobierno en Afganistán e Irak estuvo ligado a un fuerte incremento de la asistencia humanitaria oficial.

Los flujos de refugiados se prestan naturalmente para ser presentados por medio de gráficos. La Figura 10.6, elaborada para un estudio sobre desplazamiento prolongado (Crawford et al., 2015), revela la evolución del número de casos de refugiados de 10 países desde 1978 hasta 2014. El ancho de cada trazo es proporcional al número de refugiados en ese momento. Este gráfico contribuye a ilustrar el argumento de que no existe un patrón universal para el desplazamiento prolongado, sino que cada crisis sigue un patrón propio.

**Figura 10.6:** Una selección de crisis de refugiados que generaron más de 400,000 refugiados entre 1978 y el 2014



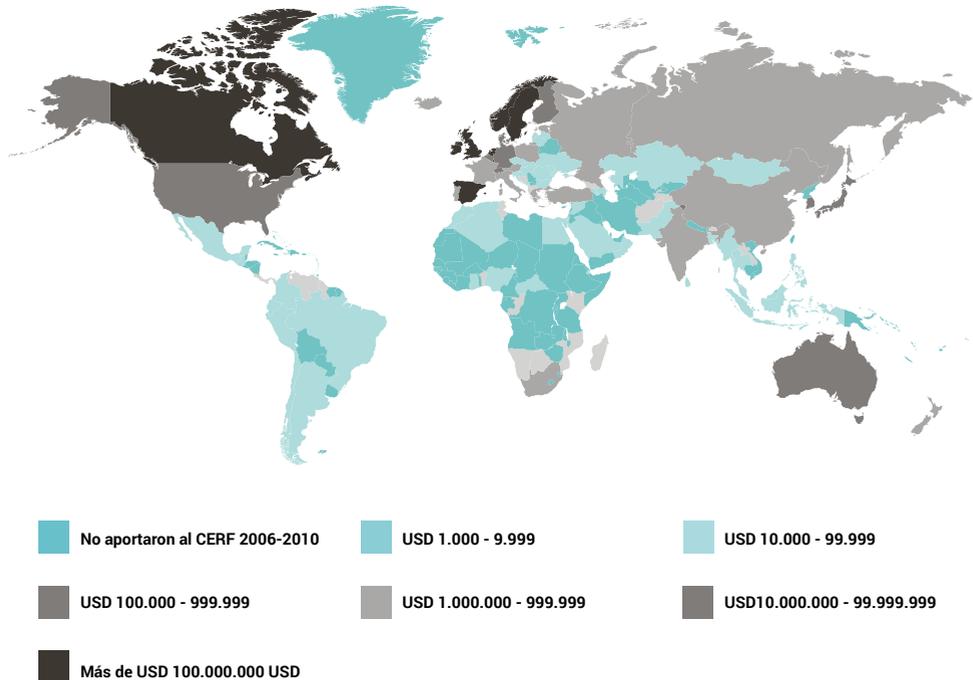
El ancho de cada trazo es proporcional al volumen de refugiados de ese país en ese año. Basado en estadísticas de ACNUR.

## Información geográfica

La información geográfica puede resumirse utilizando el programa informático *Graphical Information Systems* (GIS) para destacar aspectos clave. Se trata de paquetes de GIS baratos y de acceso libre que están disponibles pero no siempre son fáciles de manejar. El mapa a continuación muestra qué países contribuyeron al CERF durante sus primeros cinco años; los colores oscuros corresponden a los aportes más altos. Esto corrobora que el CERF goza de apoyo generalizado, incluso de los países en desarrollo.

Nuevamente, el software de análisis de contenido puede sacar a la luz la medida en que se utilizan denominaciones geográficas en un conjunto de datos en particular, y esto puede usarse para revelar su focalización geográfica.

Figura 10.7: Aportes de países al CERF entre el 2006 y el 2010

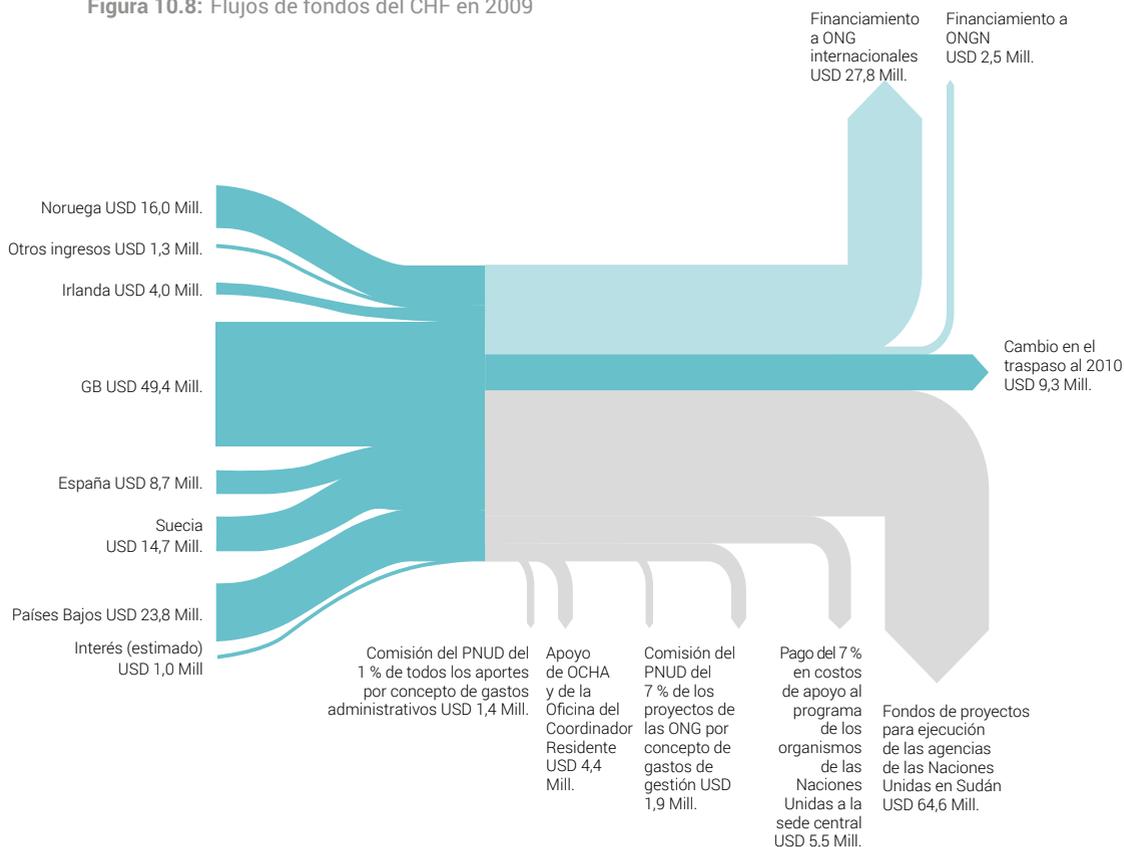


## Análisis de flujos

Los flujos complejos pueden ser más fáciles de entender si se presentan en un diagrama de Sankey. La [Figura 10.8](#), tomada de la evaluación del Fondo Común Humanitario (CHF, por sus siglas en inglés) en Sudán en 2009, reveló que se estaba canalizando relativamente poco financiamiento a las ONG nacionales, y que una porción considerable se destinaba a gastos administrativos generales.

La [Sección 16](#) abordará el análisis numérico.

**Figura 10.8:** Flujos de fondos del CHF en 2009



Fuente: [www.unsudanid.org](http://www.unsudanid.org), *gateway* del PNUD y del MDTF (Fondo del Fideicomiso Multidonante) y sección de gestión de fondos de OCHA

## Notas finales

### 9 / Planificación y gestión de su evaluación

1. Jane Davidson (2010a) incluye “optar por el consultor con la tarifa diaria más baja” en su listado irónico de nueve reglas doradas para encargar una evaluación que sea un desperdicio de dinero.
2. Para algunas organizaciones de las Naciones Unidas, por ejemplo, resulta mucho más rápido contratar a los evaluadores directamente que trabajar con una empresa consultora.

### 10 / Métodos documentales

3. Dicho estudio reveló que todas las evaluaciones noruegas de intervenciones de asistencia humanitaria incluidas en la muestra arrojaron consecuencias fortuitas.
4. Ver [www.alnap.org/current/haitilearningportal](http://www.alnap.org/current/haitilearningportal).
5. Ver [www.urban-response.org](http://www.urban-response.org).
6. Ver [www.alnap.org/resources](http://www.alnap.org/resources).
7. Ver <http://dtsearch.com>.

8 / Fase Inicial  
135 - 146

9 / Planificación y gestión de su evaluación  
147 - 171

10 / Métodos de revisión documental  
172 - 187

## Notas

## Notas

# Llevando a cabo la evaluación



# 11 / Diseños de evaluación para responder las preguntas de evaluación

Esta sección se refiere a cómo estructurar una evaluación con el fin de responder la pregunta de evaluación, lo cual a su vez puede ayudar a los evaluadores y gestores de la evaluación a considerar los diseños de evaluación posibles.

El diseño de evaluación es la lógica general detrás de la forma en que organizan su investigación para responder a una pregunta de evaluación. Algunos diseños de evaluación son más adecuados para responder a ciertos tipos de pregunta. El objetivo de esta sección es ayudarlos a considerar qué diseño o diseños de evaluación permitirán al equipo de evaluación responder de la mejor manera posible a la pregunta o preguntas de evaluación, considerando la naturaleza del programa, el contexto en que se ha llevado a cabo y las restricciones de la evaluación en términos de acceso, presupuesto, información y otros factores.

La investigación tradicional suele tener una sola hipótesis o pregunta a considerar, o bien un conjunto de hipótesis o preguntas estrechamente relacionadas. Las evaluaciones a menudo incluyen una serie de preguntas dispares, lo que puede significar que debemos utilizar más de un tipo de diseño de evaluación o un diseño apto para responder todas las preguntas pero que no es ideal para algunas preguntas.



## **Tener presente**

No confundir diseño con metodología. Diseño se refiere a la estructuración de la recolección y el análisis de información, mientras que metodología se refiere a la forma en que se recolecta la información.

Tal como se indicó en la [Sección 6, Selección de las preguntas de evaluación](#), las preguntas de evaluación determinan el diseño de evaluación, las metodologías de recolección y análisis de información y los enfoques de muestreo.

Esta sección incide también en el problema del sesgo, ya que este puede variar dependiendo del diseño, metodologías y enfoques de muestreo de la evaluación.

## 11.1 Cualitativos y cuantitativos

Algunos diseños de evaluación (por ejemplo, los estudios de caso) suelen ser asociados con enfoques cualitativos, mientras que otros (por ejemplo, los experimentos) lo son con enfoques cuantitativos.



### En profundidad: El problema con “cualitativo versus cuantitativo”

En principio, los métodos cuantitativos recogen información numérica, mientras que los métodos cualitativos recogen información no numérica. En la práctica, sin embargo, estas etiquetas son bastante más complejas, y Goetz y Mahoney (2012) sostienen que representan dos culturas de investigación distintas. Mahoney y Goetz (2006: 245) observan que “las etiquetas cuantitativo y cualitativo no llegan a capturar las diferencias reales entre las dos tradiciones. El análisis cuantitativo implica el uso de números por naturaleza, pero todos los análisis estadísticos se apoyan fuertemente también en el uso de palabras para la interpretación. Los estudios cualitativos muy frecuentemente emplean información numérica; de hecho, muchas técnicas cualitativas requieren de información cuantitativa.”

Las debilidades de ambas etiquetas han hecho que se propongan muchas otras, entre ellas *small-n*, enfoques explicativos y enfoques basados en casos para la investigación cualitativa; y *large-n* así como enfoques estadísticos, estimativos y basados en la población para la investigación cuantitativa. Mahoney y Goetz (2006: 245) mencionan 10 áreas de diferencia entre ellas. Patton (2014, Anexo 3.1) se basa en 10 factores para contrastar los enfoques cuantitativos, cualitativos y de metodología mixta.

Se trata de una discusión en cierto modo académica, pero que reviste importancia para los evaluadores humanitarios porque las dos culturas parten de teorías del conocimiento distintas. En el caso de la investigación cuantitativa, el conocimiento emana de métodos experimentales, como en las ciencias físicas. Este modelo de desarrollo de conocimiento se denomina positivismo. En el caso de la investigación cualitativa existen varias teorías que intentan explicar cómo se llega al conocimiento, entre ellas la etnografía y la teoría fundamentada. Patton (2014, Capítulo 3) menciona el uso de 16 perspectivas teóricas distintas en la investigación cualitativa.

El problema para los evaluadores es que ha habido una tendencia reciente a calificar de riguroso solamente el conocimiento generado por medios experimentales. Tanto la Asociación Americana de Evaluación (2003) como la Sociedad Europea de Evaluación (2007) han criticado este enfoque. Knox-Clarke y Darcy (2014) afirman que el uso de metodologías cualitativas o cuantitativas no genera automáticamente evidencia más robusta o más débil.



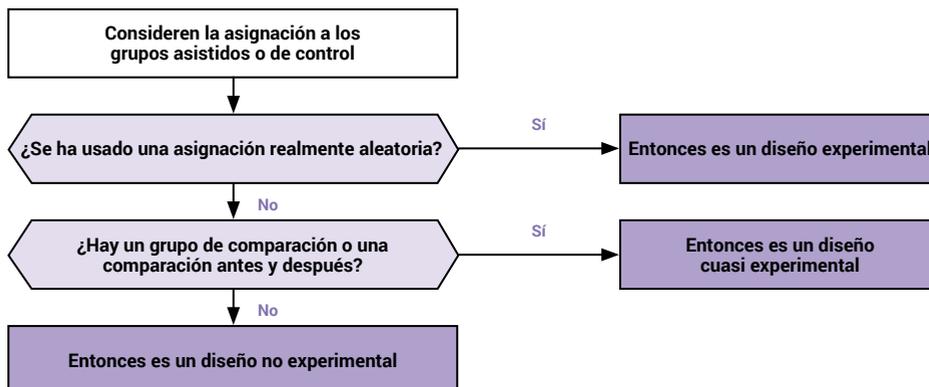
Un problema adicional para los evaluadores es que las organizaciones que encargan evaluaciones privilegian los experimentos o aplican medidas que son apropiadas para diferentes tipos de investigación. La investigación cualitativa no puede ser evaluada aplicando la terminología y los conceptos utilizados en la evaluación de la investigación cuantitativa sin un sentido crítico (Anastas, 2004: 57). El estudio médico de ACAPS distingue así los indicadores de la investigación cuantitativa de los de la investigación cualitativa: *validez interna/veracidad, validez externa/capacidad de generalización, confiabilidad/consistencia/precisión*, así como *objetividad* para la investigación cuantitativa, en oposición a *credibilidad, capacidad de transferencia, seriedad y confirmabilidad* para la investigación cualitativa (ACAPS, 2013).

## 11.2 Familias de diseño

Hay tres grandes familias de diseño de investigación (ver ejemplos más adelante):

- Diseños no experimentales sin grupos de comparación o control.  
El diseño no experimental es el tipo de diseño más común en la EAH.
- Diseños experimentales donde la asignación a los grupos asistidos o de control (grupos de tratamiento) se efectúa antes del inicio de la asistencia.
- Diseños cuasi experimentales, donde las comparaciones se efectúan ya sea del grupo asistido a través del tiempo o entre el grupo asistido y un grupo de comparación seleccionado después del inicio de la asistencia.

La categoría a la que se ajusta un diseño de evaluación depende de si se usa o no un grupo de control o de comparación. Las comparaciones pueden ser entre grupos distintos o de un mismo grupo a través del tiempo. El flujograma o diagrama de actividades que presentamos a continuación resume las diferencias entre las familias de diseño de las EAH.



Si un diseño de investigación no utiliza grupos de control ni de comparación, se trata de un diseño no experimental, que es el tipo de diseño de evaluación más común en la EAH.

Una misma evaluación puede requerir distintos diseños en secuencia para poder responder las preguntas adecuadamente. Este puede ser el caso de los diseños de metodología mixta, que pueden consistir de intervalos de métodos cuantitativos y cualitativos que utilizan diseños distintos.

Los TdR de la gran evaluación de Solidaridad Suiza 10 años después del terremoto y los tsunamis del Océano Indico originalmente disponían que se llevara a cabo una encuesta cuantitativa antes de emplear otros métodos cualitativos. Ferf y Fabbri (2014: 16) se pronunciaron a favor de revertir esta orden para “crear una entrevista que fuera lo más abierta posible, evitando las expectativas preconcebidas sobre los resultados y ofreciendo oportunidades para identificar lo inesperado tanto como fuese posible. Después de esto, la focalización de la investigación se redujo a los temas que fueron identificados como más importantes para la recuperación de los beneficiarios.”

Esto puede mejorarse aún más con un examen no experimental de seguimiento de los temas identificados en la encuesta. Este modelo, con un estudio de prefactibilidad no experimental para definir la focalización de una encuesta cuasi experimental, seguido de un examen final no experimental en profundidad de los temas emergentes, es el ciclo del modelo de investigación postulado por Remler y Ryzin (2015) y es parecido a los ejemplos de metodologías mixtas intercaladas que ofrece Bamberger (2012).

## Diseños no experimentales



### **Definición: Diseños no experimentales**

Diseños no experimentales son aquellos en los que no se realiza ninguna comparación, ya sea entre poblaciones asistidas y no asistidas o de aquellas que han recibido asistencia a través del tiempo.

La mayoría de EAH utilizan diseños no experimentales una única vez por, entre otras, las siguientes razones:

- Son los menos exigentes en términos de requisitos que cumplir;
- Son los más flexibles;
- Son relativamente económicos;
- Sirven para responder a todo tipo de preguntas de evaluación;
- Se ajustan bien a las capacidades de los evaluadores humanitarios, normalmente desarrolladas en el transcurso de diagnósticos de necesidades y planificaciones de programas.

Los diseños no experimentales pueden incluir:

- Estudios de caso;
- Revisiones de proceso;
- Revisiones de efectos;
- Diseños participativos.

En ocasiones se considera la evaluación basada en teorías como una forma de diseño no experimental, a pesar de ser más compleja. Una evaluación basada en teorías puede utilizar diversos diseños para validar la teoría del cambio subyacente, entre ellos diseños experimentales, cuasi experimentales y no experimentales.

Los estudios de caso, en que la evaluación examina una serie de unidades de análisis individuales para extraer conclusiones genéricas acerca de la intervención, probablemente sean el diseño más común utilizado en EAH. Los diseños verdaderamente participativos (en que las partes interesadas definen la evaluación) no se utilizan en EAH, tal como se explica en mayor detalle en la [Sección 14, Involucrando a la población afectada en su evaluación](#). Los evaluadores humanitarios tienen que ir más allá y no limitarse a utilizar diseños no experimentales porque:

- Otros diseños son más apropiados para responder ciertos tipos de pregunta de evaluación;
- Algunas entidades donantes están ejerciendo presión para avanzar hacia lo que en ocasiones se consideran diseños más rigurosos;
- Otros diseños pueden servir para responder algunas preguntas de evaluación a un menor costo que los diseños no experimentales (esto puede ser particularmente cierto en el caso de los diseños cuasi experimentales que utilizan información secundaria).

La [subsección sobre la selección de su diseño en la página 200](#) señala los casos en que ciertos diseños resultan más provechosos.

## Diseños cuasi experimentales



### Definición: Diseños cuasi experimentales

Diseños cuasi experimentales son aquellos que utilizan una comparación donde el grupo de comparación no es seleccionado al azar.

Los diseños cuasi experimentales engloban una amplia gama de diseños, incluidos los siguientes:

- Diseños en que el grupo asistido es comparado con un grupo de comparación (ya sea en una fecha determinada o a lo largo de un período de tiempo);
- Diseños en que un mismo grupo es comparado a través del tiempo.

En la subsección Selección de un diseño de evaluación encontrarán una serie de diseños cuasi experimentales. La principal fortaleza de estos es que pueden suministrar evidencia rigurosa de la eficacia o no de una intervención, al mismo tiempo que se evitan los problemas éticos de los diseños experimentales.

### Uso de grupos de comparación

El uso de grupos de comparación ayuda a reducir el riesgo de que cualquier cambio detectado en el grupo asistido se deba a cambios más generales en el entorno, antes que a la asistencia otorgada. Los diseños que incluyen grupos de comparación son más robustos que los diseños que no lo hacen.

No obstante, los grupos de comparación en contextos humanitarios son altamente susceptibles a la contaminación experimental debido al gran número de actores y otras redes de apoyo (incluidas las familias).

La contaminación experimental tiene lugar cuando no podemos controlar todos los aspectos de la asistencia a los grupos asistidos y a los grupos de control o comparación.

La contaminación de los experimentos y cuasi experimentos en situaciones humanitarias puede darse de muchas formas, entre ellas en circunstancias en que:

- Otras organizaciones o miembros de la familia proporcionan asistencia a los miembros del grupo de control asistido;
- Los miembros del grupo de control aprenden del grupo asistido (acerca de los beneficios de lavarse las manos antes de entrar en contacto con alimentos, por ejemplo) y adoptan la práctica mejorada.

## Problemas con los grupos de comparación – algunos ejemplos

Una evaluación de las entregas en efectivo a los refugiados de regreso en su país de origen, Burundi, había planeado comparar a los receptores y no receptores, pero muy pronto advirtió que los dos grupos no eran comparables, ya que habían regresado en épocas distintas (Haver y Tennant, 2009).

El estudio sobre niños soldados en Mozambique había planeado utilizar un grupo de comparación pero se encontró con que muchos de los niños soldados que habían seleccionado para las entrevistas no querían responder del todo sus preguntas (Boothby et al., 2006: 106). La *Guía interagencial de evaluación de programas psicosociales en crisis humanitarias* relata que esto llevó al estudio a utilizar las “normas locales” como comparación, antes que a otros niños soldados (Ager et al., 135).

Una evaluación de impacto de la ayuda alimentaria para los refugiados en Bangladesh observa que la diferencia entre los grupos asistidos y de comparación puso en riesgo la validez interna del estudio (Nielsen et al., 2012).

El estudio *La leche materna importa (Milk Matters)* en Etiopía utilizó un grupo de comparación que ya era significativamente distinto al grupo asistido, incluso antes del tratamiento (Sadler et al., 2012). Cuando existe una diferencia importante entre los dos grupos antes de la intervención, la atribución de cualquier diferencia posterior a la intervención resulta imposible.

Los grupos de comparación pueden establecerse a través de una serie de métodos. El más riguroso probablemente sea el emparejamiento por propensión, basado en técnicas estadísticas.



### **Definición: Emparejamiento por puntuación de propensión**

El emparejamiento por puntuación de propensión es una técnica estadística de coincidencia que intenta emparejar al grupo de comparación con el grupo de control mediante la selección de un grupo que tenga las mismas probabilidades de ser asistido, de acuerdo a las características del grupo.

Caliendo y Kopeinig (2005) proporcionan orientación sobre el uso del emparejamiento por puntuación de propensión en la evaluación. Algunas de las dificultades inherentes a la aplicación de este enfoque emergieron en la revisión del programa de distribución masiva de suplementos alimenticios en Níger (Grellety et al., 2012).

Otra forma de seleccionar al grupo de comparación es consultando opiniones expertas.

## Diseños experimentales

Los diseños experimentales son los más exigentes, y para aplicarlos con éxito hay que cumplir numerosas condiciones. Algunos los consideran el diseño más riguroso.



### Definición: Diseños experimentales

Diseños experimentales son aquellos donde las unidades de análisis se asignan aleatoriamente al grupo asistido o al grupo de control. Cada elemento (es decir, cada persona, familia o comunidad) tiene las mismas probabilidades de ser asignado ya sea al grupo asistido o al grupo de control.

La asignación aleatoria de la asistencia conlleva problemas éticos para la EAH, ya que la asistencia debe ser otorgada con base en las necesidades antes que al azar.

La prueba controlada aleatorizada (PCA) es el diseño experimental más común. Cuando se aplica en situaciones de EAH, la PCA en ocasiones se emplea para comparar diferentes formas de asistencia (dinero en efectivo, vales de comida o alimentos, como en la evaluación de Sandström y Tchatchua, 2010), en vez de asistencia y no asistencia, o bien en la fase de recuperación, como en la evaluación de los esfuerzos de la fase reconstrucción dirigida por la comunidad en el condado de Lofa, Liberia (Fearon et al., 2008). Algunos estudios de PCA han versado sobre el sector salud en situaciones humanitarias, por ejemplo sobre el efecto de añadir suplementos como probióticos y prebióticos a los alimentos en Malauí (Kerac et al., 2009) o de utilizar láminas de plástico tratadas con insecticida en Sierra Leona (Burns et al., 2012).

## 11.3 Selección del diseño

Idealmente, el diseño es determinado por las preguntas de evaluación únicamente. Además, no existe un diseño de evaluación perfecto. Las restricciones que plantea la falta de tiempo, presupuesto, disponibilidad de información, etc., limitan las opciones. Las opciones seleccionadas, y las razones para ello, deben constar en los informes tanto inicial como final. Esto se debe a que es posible que deban apartarse del diseño inicial una vez que estén en el terreno. La presente sección describe los elementos clave de una gama de diseños. La información entre paréntesis se refiere a la familia de diseño.

Estudio de caso (no experimental)	
 <b>Definición</b> Descripción y análisis intensivo de uno o más casos (que pueden abarcar desde personas individuales hasta Estados) con el fin de extraer conclusiones generales acerca de la intervención.	
<b>Uso en la EAH</b>  Los estudios de caso son comúnmente utilizados, generalmente en relación con familias y comunidades. Los estudios de caso nacionales pueden usarse en evaluaciones a gran escala.	
<b>Orientación</b>  Yin (2003a; 2003b) proporciona orientación sobre el uso de estudios de caso.	
<b>Fortalezas</b> <ul style="list-style-type: none"> <li>• Son efectivos para responder a preguntas sobre por qué la gente ha hecho ciertas cosas.</li> <li>• Son efectivos para examinar circunstancias inusuales o intervenciones complejas.</li> <li>• Son efectivos para validar una teoría.</li> <li>• Suministran rica información sobre los casos.</li> <li>• Son efectivos para desarrollar teorías.</li> <li>• Se ajustan bien con la EAH.</li> </ul>	<b>Debilidades</b> <ul style="list-style-type: none"> <li>• Pueden ser difíciles de generalizar.</li> <li>• Los casos suelen ser seleccionados intencionalmente, tornando más difícil la generalización a partir de los hallazgos.</li> <li>• Menos útiles para estudios sobre la atribución.</li> </ul>
<b>Ejemplos</b>  La evaluación de 2015 de la asistencia humanitaria danesa utilizó dos estudios de caso en profundidad a nivel de campo (Sudán del Sur y crisis de Siria) y un estudio documental más limitado (Afganistán) (Mowjee et al., 2015).	

Revisión de proceso (no experimental)	
 <b>Definición</b> Una revisión de proceso compara la forma en que funcionan los procesos con la forma en que se tenía previsto que funcionarían.	
<b>Uso en la EAH</b>	
Las revisiones de proceso se usan para verificar si los procesos se ajustan a normas más generales, como las normas mínimas del Proyecto Esfera o las normas de la propia organización.	
<b>Orientación</b>	
La mayor parte de la orientación sobre este diseño se refiere a revisiones de proceso empresariales. No se ha encontrado ninguna guía que tratara específicamente sobre revisiones de proceso humanitarias.	
<b>Fortalezas</b>	<b>Debilidades</b>
<ul style="list-style-type: none"> <li>• Son efectivos para responder preguntas normativas.</li> <li>• Son efectivos para responder preguntas sobre los supuestos de la planificación.</li> <li>• Son efectivos para responder preguntas sobre la eficiencia de los procesos.</li> <li>• Pueden realizarse en cualquier momento de una intervención.</li> </ul>	<ul style="list-style-type: none"> <li>• Suelen prestar escasa atención al impacto.</li> <li>• Son efectivos para algunas preguntas solamente.</li> <li>• Existe el supuesto implícito de que un proceso adecuado garantizará resultados.</li> </ul>
<b>Ejemplos</b>	
La revisión de evaluaciones participativas realizada por ACNUR en 2013 es un ejemplo de una revisión de proceso. La revisión de proceso del Fondo Común Humanitario en Somalia es otro (Willitts-King et al., 2012).	

11 / Diseños de evaluación para responder las preguntas de evaluación 192 - 213

12 / Muestreo 214- 230

13 / Metodologías de campo 231 - 267

14 / Involucrando a la población afectada en su evaluación 268- 282

15 / Acceso restringido 283 - 292

16 / Análisis 293 - 321

## Revisión de efectos (no experimental)



### Definición

Una revisión de efectos compara los efectos constatados con los efectos programados.

### Uso en la EAH

Uso formal limitado, pero muchos informes de EAH se enfocan en los efectos antes que en el impacto.

#### Fortalezas

- Evitan la a menudo difícil tarea de evaluar la contribución o la atribución de los impactos.
- Se ajustan bien a los documentos de planificación de programas.

#### Debilidades

- No inciden en el impacto. Si bien es posible que se hayan logrado efectos, estos no necesariamente generan impactos positivos.
- No indican por qué ocurren las cosas.

### Ejemplos

No existen ejemplos formales de este tipo de diseño en EAH, pero la evaluación de Oxfam GB de la respuesta a la sequía en Kenia (Turnbull, 2012) alude a una revisión anterior de efectos de los programas de transferencias en efectivo.

## Diseño participativo (no experimental)



### Definición

El diseño participativo promueve la participación de todas las partes interesadas en todas las fases de la evaluación, desde la planificación inicial hasta la puesta en práctica de las recomendaciones. (Ver también evaluación participativa en la [Sección 4.](#))

### Uso en la EAH

No se hallaron ejemplos reales de este tipo de diseño en la EAH. Si bien es cierto que muchas evaluaciones se esfuerzan por ser participativas, no se encontró ningún ejemplo de una evaluación en la que todo el proceso estuviera controlado por las partes interesadas. Este enfoque es utilizado principalmente en el sector de la filantropía estadounidense. Los métodos participativos se utilizan con frecuencia en la evaluación (Catley, 2009; 2014), pero una evaluación es clasificada como participativa solo si las partes interesadas definen las preguntas de evaluación.

### Orientación

Existe una breve nota de orientación de USAID sobre la evaluación participativa (USAID, 1996), así como un manual de CRS más detallado (Aubel, 1999). Un buen ejemplo de una evaluación participativa es la evaluación del programa de transversalización del enfoque de edad, género y diversidad de ACNUR en Colombia (Mendoza y Thomas, 2009).

#### Fortalezas

- Empodera a los participantes.
- Promueve su utilización.
- Promueve alianzas.

#### Debilidades

- Los resultados son difíciles de predecir.
- Sujetas a los sesgos existentes entre las partes interesadas (para continuar con el financiamiento, por ejemplo).

Series temporales interrumpidas (cuasi experimental)	
 <p><b>Definición</b> El diseño basado en series temporales interrumpidas proporciona una estimación del impacto de una intervención, al examinar una serie temporal de información antes y después de una intervención.</p>	
<p><b>Uso en la EAH</b></p> <p>Rara vez se utilizan en la actualidad.</p>	
<p><b>Orientación</b></p> <p>Cochrane ofrece un ejemplo elaborado corto de un diseño de series temporales interrumpidas (EPOC, 2013). Glass (1997) ofrece numerosos ejemplos adicionales del uso de este tipo de diseño. Gilmour et al. (2006) dan un ejemplo donde la fecha exacta de la intervención no está clara. Wagner y Ross-Degnan (2002) ofrecen orientación sobre el uso del análisis de regresión segmentada o series temporales interrumpidas.</p>	
Fortalezas	Debilidades
<ul style="list-style-type: none"> <li>Se trata de un diseño económico, dado que tiende a utilizar información secundaria ya existente.</li> </ul>	<ul style="list-style-type: none"> <li>Sujeto a contaminación, especialmente en razón de efectos temporales como la maduración (cambios conductuales a través del tiempo), etc.</li> <li>Requiere de un control cuidadoso para eliminar causas alternativas.</li> <li>Requiere de competencias estadísticas fuertes.</li> </ul>
<p><b>Ejemplos</b></p> <p>No parece haber ninguna EAH que utilice un diseño de series temporales interrumpidas, pero dicho diseño ha sido usado en estudios no evaluativos de situaciones humanitarias, tales como el suicidio tras el terremoto de 1999 en Taiwán (Xang et al., 2005) o la neumonía tras el accidente nuclear de Fukushima (Daito et al., 2013).</p>	

## Diferencias en diferencias (cuasi experimental)



### Definición

Este diseño estima los efectos de la asistencia comparando el cambio promedio a través del tiempo en el resultado de interés entre el grupo asistido y un grupo de comparación.

### Uso en la EAH

No existen ejemplos de uso de este tipo de diseño en la EAH. Algunas evaluaciones aluden a la metodología pero no la aplican rigurosamente.

### Fortalezas

- Puede ser económico si se utiliza información secundaria.
- Cuando se observan grandes cambios, arroja imágenes contundentes de la diferencia.
- Es mucho más intuitivo que muchas otras técnicas estadísticas.

### Debilidades

- Requiere de competencias estadísticas fuertes.
- El grupo de comparación debe ser lo suficientemente parecido.

## Comparación antes y después (cuasi experimental o experimental)



### Definición

Una comparación antes y después compara la situación de un grupo antes y después de la intervención.

### Uso en la EAH

Escasamente utilizado en la EAH.

### Fortalezas

- Demuestra inequívocamente que se ha producido un cambio.
- Revela la magnitud del cambio (el nivel de los efectos).
- Puede ser económico si existe información secundaria disponible.

### Debilidades

- No hay un grupo de comparación, de modo que cualquiera de los cambios observados puede obedecer a otros factores, por ejemplo, a la economía externa, etc.

### Ejemplos

Un buen ejemplo del uso de información de línea de base para un estudio antes y después fue el estudio de los sobrevivientes de las inundaciones de Bangladesh por Durkin et al. (1993). Utilizando información de un estudio no relacionado realizado seis meses antes, los autores compararon el comportamiento de los niños y niñas evaluados en ese momento, con su conducta después de las inundaciones, sugiriendo que mostraban ciertos signos de estrés posttraumático.

## Grupo de comparación (cuasi experimental o no experimental)



### Definición

Los diseños con grupo de comparación comparan al grupo asistido con un grupo de comparación seleccionado.

### Uso en la EAH

En ocasiones encontramos grupos de comparación débiles en la EAH. Es posible establecer grupos de comparación más robustos.

### Fortalezas

- Los grupos de comparación disminuyen el riesgo de confundir los cambios en el contexto con el impacto de la intervención.
- Los grupos de comparación pueden aumentar el poder de convicción de los hallazgos.

### Debilidades

- Es muy difícil evitar la contaminación en contextos humanitarios.
- Se requieren fuertes competencias estadísticas y buena información sobre el grupo asistido para usar técnicas como emparejamiento por propensión.

## Diseño de regresión discontinua (cuasi experimental)



### Definición

El diseño de regresión discontinua compara las líneas de regresión de la variable de interés con la puntuación en la que se basó la intervención.

### Uso en la EAH

No se halló ningún ejemplo real de un diseño de regresión discontinua, aunque algunos estudios de tratamiento discontinuo se describen a sí mismos como diseños de regresión discontinua.

### Orientación

Jacob et al. (2012) proporcionan orientación sobre este diseño, al igual que la plataforma interactiva *Better Evaluation* y Khander et al. (2010: ch.7).

### Fortalezas

- Se ajustan bien a la acción humanitaria, dado que el grupo asistido y el grupo de comparación son seleccionados con base en las necesidades.
- Pueden generar imágenes convincentes.
- Más robusto que una simple comparación por debajo y por encima del punto de corte.

### Debilidades

- Requiere de competencias estadísticas fuertes.
- Riesgo de contaminación si no se aplican estrictamente los criterios para el punto de corte.
- Los resultados solo son generalizables alrededor del punto de corte.

## Comparación por tratamiento discontinuo (cuasi experimental)



### Definición

El diseño de tratamiento discontinuo compara al grupo justo por debajo del punto de corte para la asistencia con el grupo justo por encima de éste.

### Uso en la EAH

Cierto uso, aunque algunos estudios son erróneamente etiquetados como diseños de regresión discontinua.

### Fortalezas

- No plantean objeciones éticas.
- Se ajustan bien a la acción humanitaria, dado que el grupo asistido y el grupo de comparación son seleccionados con base en las necesidades.
- No necesariamente requieren competencias estadísticas fuertes.

### Debilidades

- Riesgo de contaminación si no se aplican estrictamente los criterios para el punto de corte.
- Los resultados solo son generalizables alrededor del punto de corte.

### Ejemplos

La revisión de Lehmann y Masterson (2014) del enfoque de transferencias en efectivo en Líbano utiliza este diseño.

## Diseño longitudinal (cuasi experimental o no experimental)



### Definición

Los estudios longitudinales hacen mediciones repetidas de una misma población a través de los años.

### Uso en la EAH

Se usa poco, salvo para analizar la información secundaria de dichos estudios. El uso más común se da en estudios de cohortes para el sector salud en los países desarrollados.

### Fortalezas

- Indicación robusta de las tendencias graduales a través del tiempo

### Debilidades

- Estos estudios son caros de realizar.
- Identifican los cambios y las tendencias, mas no las razones subyacentes.

### Ejemplos

Khoo (2007) se basó en las encuestas longitudinales de inmigrantes a Australia para investigar la participación de los inmigrantes humanitarios en el sistema de salud y en la economía de ese país, empleando técnicas de regresión logística y dos cohortes comparables.

## Prueba controlada aleatorizada (PCA) (diseño experimental)

 **Definición**  
 La prueba controlada aleatorizada (PCA) compara dos grupos seleccionados aleatoriamente, a uno de los cuales se le otorga asistencia, mientras que el otro (el grupo de control) no recibe ninguna.

### Uso en la EAH

Existen muy pocos ejemplos en razón de las objeciones éticas, el costo y la complejidad. En ocasiones se han utilizado PCA en la fase de recuperación o para comparar diferentes módulos de asistencia, en vez de utilizar un grupo de control que no reciba ninguna asistencia.

Fortalezas	Debilidades
<ul style="list-style-type: none"> <li>• Algunos consideran que es el diseño más riguroso.</li> <li>• Es efectivo para responder preguntas causales acerca de la atribución.</li> <li>• Alto nivel de confianza en que la respuesta constituye un reflejo acertado de la verdad en el contexto.</li> <li>• La asignación aleatoria soluciona en gran medida los problemas de sesgos en la selección.</li> </ul>	<ul style="list-style-type: none"> <li>• Un gran número de objeciones éticas (entre ellas asignar aleatoriamente la asistencia, en vez de hacerlo con base en las necesidades).</li> <li>• Es el diseño más costoso.</li> <li>• Es el diseño menos flexible.</li> <li>• Exigentes prerrequisitos.</li> <li>• Apropiado únicamente para tipos específicos de preguntas.</li> <li>• Rara vez es posible mantener a los grupos en la ignorancia respecto de su pertenencia al grupo asistido o al grupo de control – generando contaminación.</li> <li>• La contaminación es difícil de controlar (por ejemplo, el grupo de control puede recibir asistencia similar de otra fuente).</li> <li>• Un marco muestral (un listado que más o menos incluya a toda la población) puede ser difícil de obtener en un contexto humanitario.</li> <li>• Necesidad de seleccionar aleatoriamente a los grupos asistidos y no asistidos antes de la intervención.</li> <li>• Capacidad de generalización limitada debido a que los resultados sirven solamente para este único experimento.</li> </ul>

### Ejemplos

La evaluación de impacto realizada por Fearon et al. en 2008 sobre los esfuerzos de reconstrucción dirigidos por la comunidad en el condado de Lofa empleó una PCA.

## Experimento natural (cuasi experimental)



### Definición

Los diseños experimentales naturales efectúan una comparación entre un grupo asistido y un grupo similar que no ha sido asistido por azar.

### Uso en la EAH

Ocasionalmente utilizado cuando los investigadores advierten que debido a las circunstancias se ha producido un experimento natural.

### Orientación

Los experimentos naturales se describen en el Capítulo 15 de Remler y Ryzin (2014).<sup>1</sup>

### Fortalezas

- Son efectivos para abordar preguntas causales sobre la atribución.
- Alto nivel de confianza en que la respuesta constituye un reflejo acertado de la verdad en el contexto.
- No tienen las consideraciones éticas asociadas a los experimentos deliberados.
- Es un diseño económico.

### Debilidades

- Reparos en el sentido que el grupo asistido y el grupo de comparación pueden ser sutilmente diferentes, generando por lo tanto sesgos en el resultado.
- El investigador no puede programarlos, limitándose a descubrirlos si han tenido lugar.

### Ejemplos

Un ejemplo de un experimento natural es el estudio sobre la afectación de los niños y niñas en Nicaragua a consecuencia del huracán Mitch en 1998 (Baez y Santos, 2007). Otros ejemplos son los estudios sobre: el impacto a largo plazo de la hambruna en los sobrevivientes (Meng y Qian, 2009), el impacto del desabastecimiento en la nutrición infantil (Boonstra et al., 2001), el impacto de los refugiados en la economía local (Maystadt, 2011), los riesgos asumidos después de un desastre (Cameron y Shah, 2012; Page et al., 2012) y las restricciones de seguridad en el apoyo a la violencia (Longo et al., 2014).

Un ejemplo más simple de un experimento natural se dio después del terremoto de 2006 en Yogyakarta. Algunas jurisdicciones municipales locales ya se encontraban en estado de alerta en previsión de una posible erupción del Monte Merapi y por lo tanto contaban con una infraestructura de preparación y gestión de desastres. El Fritz Institute encontró altos niveles de satisfacción con la respuesta a la emergencia en dichas zonas (Bliss y Campbell, 2007).

## 11.4 Sesgo

La selección del diseño, metodologías y enfoques de muestreo influye en los potenciales tipos de sesgo. El sesgo constituye una amenaza para la exactitud de los hallazgos de una evaluación (Knox-Clarke y Darcy, 2014: 15).

Existe una diferencia entre los sesgos en las metodologías cualitativas y cuantitativas. En las metodologías cuantitativas se acepta deliberadamente el sesgo – las muestras se seleccionan intencionalmente, y los investigadores deben reflexionar acerca del impacto tanto de sus propios sesgos como de aquellos generados por el proceso de investigación. Sin embargo, incluso a pesar de ello el sesgo comporta una amenaza significativa para la validez interna de las metodologías cuantitativas.

Validez interna es el grado en el que se justifican las conclusiones acerca de la causa y el efecto. Los sesgos pueden surgir a consecuencia de los prejuicios del evaluador y/o del diseño de la evaluación. Nadie es inmune a los sesgos, pero un sesgo que tiene su origen en el diseño de la evaluación es insidioso, ya que puede no ser evidente.

### Sesgo del diseño de evaluación

El sesgo del diseño se produce por errores sistemáticos en el diseño de la evaluación. Sackett (1979) identificó 35 tipos distintos de sesgo susceptibles de afectar a una investigación analítica, entre ellos varias formas de sesgo de selección y medición.<sup>2</sup>

El sesgo de selección ocurre cuando elementos de la muestra seleccionada se encuentran sesgados de alguna manera. Un ejemplo sería si se hubieran utilizado listas de distribución de beneficiarios como el marco muestral del cual se seleccionó la muestra. Esto sesgaría la muestra si un número importante de personas no hubiera recibido asistencia. El uso de cualquier metodología que no sea una selección auténticamente aleatoria de la muestra a partir del marco muestral, también supone un riesgo de sesgo.

El sesgo de medición ocurre cuando la medición se encuentra sesgada de alguna manera. El sesgo de medición puede ocurrir por diversas razones. Por ejemplo, si una balanza colgante suspendida de un travesaño no se encuentra debidamente nivelada, todas las mediciones serán inexactas. Otro ejemplo es si los entrevistados de una encuesta exageran el número de personas que viven en el hogar, con la esperanza de obtener más asistencia.

En los diseños experimentales, la selección aleatoria ayuda a prevenir los sesgos de selección, y a ocultar si una persona pertenece al grupo asistido o al grupo de control, tanto frente a los integrantes del grupo como al personal que trabaja con ellos (figura conocida como doble desconocimiento). Esto ayuda a eliminar los sesgos.<sup>2</sup> Los estudios cuasi experimentales están sujetos a potenciales sesgos tanto de selección como de medición, pero estos pueden ser controlados en cierta medida a través de mediciones estadísticas.

El muestreo intencional es deliberadamente sesgado porque se seleccionan los casos más apropiados para las preguntas de evaluación. Los diseños no experimentales, sin embargo, están sujetos a un riesgo de sesgo de selección no intencional debido a la exclusión de ciertos grupos (los jóvenes o los hogares más alejados de la carretera, por ejemplo).

Tanto en los diseños cuasi experimentales como en los no experimentales, un enfoque de muestreo transparente y riguroso, basado en criterios explícitos establecidos con anticipación, puede ayudar a controlar el sesgo del diseño de evaluación. Incluir un plan de muestreo en el informe inicial sería de utilidad en este sentido.

### **Evaluando, fuente y sesgo de publicación**

Las organizaciones tienden a evaluar lo que perciben como programas exitosos antes que fallidos. En la sesión de la Asociación Americana de Evaluación de 2009 sobre alternativas al contrafactual convencional se mencionó que “muchas evaluaciones tienen un sesgo positivo sistemático, ya que la mayor parte de la información se recoge de personas que se han beneficiado del proyecto” (Bamberger et al., 2009: 2). La guía del Grupo de Evaluación de las Naciones Unidas anota que, dado que las organizaciones ejecutoras se benefician de los proyectos, esto contribuye a un sesgo positivo (UNEG, 2011: 22).

El informe “Estado del sistema humanitario” de 2015 (2015: 29) establece que más del 50 por ciento de las evaluaciones califican el desempeño del sujeto de evaluación como bueno. La mayoría de las evaluaciones son comisionadas por organizaciones o donantes que tienen un interés creado en los resultados, de modo tal que es muy posible que la estructura de incentivos se incline más hacia los hallazgos positivos, incluso cuando se contrata a consultores externos para llevar a cabo la evaluación. Una breve lectura de los resultados nos llevó a concluir que un desempeño “bueno” también equivale a menudo a un desempeño “adecuado”.

Desde luego, en lugar de decir “el 50 por ciento de las evaluaciones” probablemente deberíamos haber dicho “el 50 por ciento de las evaluaciones publicadas”, ya que las evaluaciones también están sujetas a sesgos de publicación, lo cual reduce las probabilidades de que las evaluaciones críticas sean publicadas.

Cuadro 11.1: Ejemplo de discusiones de grupos focales y razones

Grupo	Grupo asistido	Grupo de comparación	Razones
Hombres adultos	3 grupos focales	3 grupos focales	Investigaciones previas demuestran que los hombres y las mujeres valoran distintos elementos de la asistencia de manera diferente.
Mujeres adultas	3 grupos focales	3 grupos focales	
Jóvenes	2 grupos focales	2 grupos focales	Un grupo de hombres y uno de mujeres en cada caso. Algunos informes revelan preocupación por el impacto de la asistencia en los jóvenes.
Personas con diversidad funcional	1 grupo focal	1 grupo focal	Un solo grupo debido al reducido número de personas con diversidad funcional
Adultos mayores	1 grupo focal	1 grupo focal	Un solo grupo debido al reducido número de adultos mayores

### Sesgos de recolección de información

La evaluación de la respuesta a los ciclones tropicales que azotaron el Caribe en 2004 (Gamarra et al., 2005: 3) incluyó una sección reducida sobre la forma en que el equipo abordó los sesgos, entre ellos el sesgo institucional, el sesgo de ubicación y el sesgo de memoria.

Entre las medidas de control de los sesgos se encuentran la triangulación y otros enfoques, entre ellos los siguientes:

- Entrevistar a personas que vivan lejos de la carretera o que hayan sido entrevistadas con frecuencia. Incluir la mayor cantidad posible de entrevistados que no pertenezcan a la organización y controlar los sesgos de memoria cotejando la información con informes de situación (Gubbls y Bouquest, 2013: 18).
- Entrevistar a la población afectada sin que haya personal de la organización presente. Evitar hacer preguntas directas sobre el desempeño de la organización, utilizando evidencia indirecta sobre la relación, la percepción del proyecto y las percepciones de apropiación (Cosgrave et al., 2010: 14).
- Garantizar el equilibrio de género entre los entrevistados con el fin de limitar los sesgos de género (Barakat et al., 2006: 139).
- Cotejar la información de manera rutinaria con la gama más amplia posible de personas (Iniciativas Humanitarias et al., 2001: Vol. 6.1).

- Contratar un consultor externo como jefe del equipo, con un equipo compuesto por personal de las organizaciones evaluadas (Reed et al., 2005: 1-2). Evitar que personas con experiencia previa en una organización en particular lideren las entrevistas a dicha organización (Cosgrave et al., 2008: 21).
- Utilizar herramientas colaborativas para compartir los hallazgos emergentes y revisarlos con los miembros del equipo. Identificar abiertamente cualquier potencial sesgo del equipo de evaluación (Ternström et al., 2007: 90-91).
- Analizar los resultados país por país para evitar los sesgos atribuibles a la sobrerrepresentación de algunos países en una encuesta en línea (Steets et al., 2013: 15).
- No informar a los encuestadores sobre el propósito del estatus de tratamiento de una comunidad, y no informar a la población afectada sobre el propósito de la encuesta (Beath et al., 2013: 46).

### Sesgo de género y edad

Los sesgos de género y edad pueden aparecer durante la recolección de información y proyectar una imagen errónea de lo que está sucediendo.

UNEG ha publicado una guía sobre la incorporación de los derechos humanos y la igualdad de género en las evaluaciones (2014).

### Prevención de sesgos de género y edad en la recolección de información de campo

Las medidas que se señalan a continuación ayudarán a asegurar que una evaluación se lleve a cabo con enfoque de género y edad y arroje información apta para ser desagregada por sexo y edad:

- Cerciorarse de que el equipo que realizará el trabajo de campo esté formado por igual número de mujeres y hombres. Las entrevistadoras y traductoras de sexo femenino suelen tener un mejor acceso a las mujeres y niñas de la población afectada.
- Cerciorarse de que entre los informantes clave se encuentren mujeres que tengan un buen conocimiento de la comunidad y de las necesidades particulares de las mujeres y si estas han sido asistidas y cómo. Las informantes clave podrían incluir maestras y enfermeras, así como dirigentes femeninas del mercado. Cerciorarse de que algunos informantes clave sepan también sobre las necesidades de los niños y niñas, los jóvenes y los adultos mayores.
- Cerciorarse de que al menos la mitad de los participantes en los grupos focales y entrevistas grupales sean mujeres. Idealmente, las mujeres deben ser entrevistadas separadamente de los hombres, quienes tienden a dominar la discusión cuando se trata de grupos mixtos.





- Llevar a cabo los grupos focales y entrevistas grupales por grupos de edad, por ejemplo con niños, niñas, adolescentes y adultos mayores. Los adultos tienden a dominar las discusiones cuando se trata de grupos mixtos.
- Anotar en un registro el género de los informantes clave y otros entrevistados con el fin de detectar posibles sesgos de género en la información recogida en las entrevistas.

Fuente: Basado parcialmente en Mazurana et al. (2011)

### Sesgo de los evaluadores

Los sesgos de los evaluadores pueden incluir su antipatía hacia una organización o un enfoque programático en particular o la tentación de reprimir hallazgos altamente críticos por temor a perder futuros contratos con el cliente. Dichos sesgos pueden regir para todas las modalidades de diseños de investigación, pero las metodologías no experimentales suponen el riesgo más alto. Estos sesgos pueden ser controlados hasta cierto punto utilizando la triangulación entre varios investigadores, es decir, comparando los hallazgos de diferentes investigadores.

Entre los enfoques que pueden contribuir a controlar los sesgos de los evaluadores se cuentan los siguientes:

- Contratación de evaluadores que sean conscientes de sus propios sesgos y traten de tenerlos en cuenta;
- Toma de decisiones transparentes con relación al muestreo;
- Estrategias transparentes de recolección y análisis de información;
- Vínculos claros entre la evidencia, los hallazgos y las conclusiones;
- Aseguramiento de la calidad externo;
- Triangulación de los hallazgos entre evaluadores;
- Encargar a más de un evaluador que codifique la misma información y comparar los resultados en busca de sesgos sistemáticos.



#### Consejo útil

Sean transparentes con respecto a los sesgos. Incluyan una declaración en el informe evaluación indicando cuáles son los potenciales sesgos del equipo de evaluación e identifiquen qué medidas han tomado para controlar tanto esta como otras modalidades de sesgo.

# 12 / Muestreo

En esta sección se analiza el muestreo. En ocasiones es posible investigar cada caso individual, pero con frecuencia hay demasiados casos o demasiados posibles informantes como para consultar con todos ellos. En vez de eso, tomamos una muestra.

**Definición: Muestreo**

Selección de un subconjunto poblacional para incluirlo en un estudio, en vez de toda la población.

Prácticamente todas las evaluaciones utilizan el muestreo. La presente sección explica cómo hasta la evaluación más pequeña puede tomar buenas decisiones con respecto al muestreo con la finalidad de mejorar la calidad de la evaluación.

En su expresión más simple, la muestra se toma por conveniencia de la población afectada, o de informantes o instancias clave. En su expresión más compleja, las muestras pueden tomarse aleatoriamente de diferentes estratos de la sociedad. Se esperaría que una evaluación pequeña de buena calidad utilizara:

- Un enfoque de muestreo no aleatorio para la recolección de información cualitativa;
- Un enfoque de muestreo aleatorio para la recolección de información cuantitativa.

**Definición: Muestreo no aleatorio**

El muestreo no aleatorio elige la muestra en función de alguna característica de la muestra.

En las muestras no aleatorias, la posibilidad de que un miembro dado de una población sea seleccionado depende de las características del mismo: es decir, no todos tienen la misma posibilidad. Por eso, las muestras no aleatorias no son representativas de la población en su conjunto.



### **Definición: Muestreo aleatorio**

El muestreo aleatorio extrae una muestra de la población en que cada miembro tiene la misma posibilidad de ser seleccionado.

Dado que en una muestra aleatoria cada miembro de la población tiene la misma posibilidad de ser seleccionado, dichas muestras son consideradas representativas.

Esta sección analiza los factores que los evaluadores deben tener en cuenta a la hora de diseñar su estrategia de muestreo. Los evaluadores deben contar con una estrategia de muestreo explícita, diseñada para responder las preguntas de evaluación de manera verosímil.

## 12.1 Problemas muestrales comunes en la EAH

Los siguientes problemas muestrales se observan con frecuencia en la EAH:

- Dependencia excesiva de la disponibilidad o del muestreo por conveniencia en investigaciones de pocos casos o *small-n*. Se trata de un enfoque débil. Morra y Rist (2009: 363) sostienen que “las muestras por conveniencia constituyen una forma sumamente débil de inferir cualquier tipo de evidencia o patrón a partir de la información.” El muestreo por conveniencia de la población afectada es muy engañoso. Si la asistencia hubiera sido otorgada con base en un criterio de conveniencia similar al de la muestra, la sugerencia sería que todo el mundo habría sido asistido.
- Tamaño de la muestra aleatoria insuficiente como para hacer generalizaciones estadísticamente válidas a partir de la muestra.
- Uso inapropiado del muestreo aleatorio en investigaciones de pocos casos o *small-n*: Collier y Mahoney (1996: 57) señalan que, en las muestras *small-n*, “la estrategia de evitar los sesgos de selección por medio de un muestreo aleatorio puede crear igual número de problemas que los que resuelve.”
- La omisión de explicitar el enfoque de muestreo en los informes de evaluación, especialmente cuando se usan metodologías *small-n*.

Daniel (2013) ofrece una introducción en profundidad del muestreo probabilístico y no probabilístico. Un documento simple y básico sobre el muestreo es el folleto *Practical Guide to Sampling (Guía práctica de muestreo)*, de la Oficina Nacional de Auditoría del Reino Unido (2000).

## 12.2 Muestreo no aleatorio

El muestreo aleatorio a menudo resulta inapropiado para los enfoques cualitativos que son los que predominan en gran parte de la EAH. La guía de muestreo de Daniel (2013, Capítulo 3) sugiere que un muestreo no aleatorio puede ser apropiado en los siguientes casos:

- Si la investigación es de índole exploratoria;
- Si debe tomarse una decisión rápida;
- Si la acción debe focalizarse en personas específicas;
- Si la intención es proveer ejemplos ilustrativos;
- Si el acceso es difícil o la población se encuentra muy dispersa;
- Si el tiempo y el dinero son limitados pero hay acceso a personal calificado y altamente capacitado;
- Si no existe un marco muestral;
- Si se están usando métodos cualitativos;
- Si deben usarse procedimientos operativos fáciles;
- Si el tamaño de la población objetivo es reducido.

La dificultad de acceso es común en la EAH. En una evaluación en la República Democrática del Congo (RDC), los evaluadores comentaron lo siguiente:

El contexto dinámico y la logística en la región oriental de la RDC significaron que era imposible seleccionar aleatoriamente las ubicaciones por adelantado, especialmente en Kivu del Norte y del Sur, que experimentaron conflictos y desplazamientos frecuentes durante gran parte de 2012. Al mismo tiempo, las ubicaciones se seleccionaron oportunamente a corto plazo, de modo tal que los miembros del equipo pudieran observar las actividades en curso. (Backer et al., 2013: 12)

En las investigaciones de pocos casos o *small-n* rara vez se requieren muestras representativas o se necesita entrevistar a un director que haya permanecido en el lugar el tiempo promedio, pero resulta útil entrevistar al director que más tiempo haya estado que pueda proveer la mayor cantidad de información. Si queremos saber qué factores contribuyeron al éxito de un programa de microcrédito y estamos utilizando un enfoque cualitativo, sesgamos nuestra muestra si entrevistamos a los prestatarios más exitosos y les preguntamos cuáles fueron las razones de su éxito. Podemos entrevistar a los prestatarios menos exitosos también, para averiguar por qué no tuvieron éxito. Esto significa que no podemos generalizar las razones y hacerlas extensivas a toda la población, pero sí podemos decir algo acerca de los factores de éxito.

El muestreo intencional probablemente sea el tipo de muestreo más robusto para estudios de tipo *small-n* porque los integrantes de la muestra se escogen deliberadamente en razón del conocimiento que pueden aportar a la investigación. Es por eso que el criterio que se utiliza para definir el tamaño de la muestra en los muestreos no aleatorios es la saturación de información.

**Definición: Saturación de información**

La saturación de información se produce cuando los casos nuevos dejan de aportar conocimiento nuevo.

El riesgo que entrañan las muestras *small-n* no es que no sean representativas, sino que no incluyen todas las categorías importantes.

**Tipos de estrategia muestral no aleatoria**

Daniel (2013) dividió el muestreo no probabilístico en cuatro categorías – por disponibilidad, intencional, por cuotas y con muestreo asistido por los propios encuestados. La estrategia de muestreo apropiada depende de la metodología y del tipo de pregunta que el equipo esté intentando responder.

**Muestreo por disponibilidad****Definición: Muestreo por disponibilidad**

El muestreo por disponibilidad es un procedimiento muestral donde la selección se basa en la disponibilidad, la conveniencia del investigador o la autoselección.

Este tipo de muestreo se conoce comúnmente como muestreo por conveniencia, y Daniel reserva esta denominación para un subtipo. Mientras que Patton afirma que “el muestreo por conveniencia no es ni estratégico ni intencional. Es ocioso y en gran medida inservible” (2014: Module 39), Daniel asegura que es “el procedimiento muestral más utilizado en la investigación”. El muestreo por disponibilidad tiene la ventaja de ser económico, simple y conveniente. Sus principales debilidades son que es la metodología no aleatoria menos confiable, conlleva una sobrerrepresentación de aquellos que están más disponibles y suele subestimar la variabilidad de la población. Debe usarse solo cuando no existan alternativas viables.

La peor combinación posible consiste en emplear el muestreo por conveniencia cuando la asistencia también ha sido provista con base en la conveniencia.



## Muestreo intencional

### **Definición: Muestreo intencional**

Un muestreo intencional selecciona la muestra de manera intencionada, de modo tal que los elementos seleccionados puedan proporcionar la mayor cantidad de información para el estudio.

Las muestras intencionales pueden seleccionarse con base en los siguientes criterios:

- Reputación y experiencia. El equipo puede preferir hablar con los directores humanitarios más experimentados, o con la población afectada que ha estado más expuesta al proyecto o programa que está siendo evaluado.
- Si se estima que están en condiciones óptimas para proporcionar la información requerida para responder a las preguntas de evaluación. Por ejemplo, si una pregunta de evaluación indaga si los grupos vulnerables fueron tratados con dignidad, los miembros de los grupos vulnerables estarán en la mejor posición para responder.
- Si se estima que son los que mejor pueden demostrar un componente determinado de la teoría del cambio (TdC). Estos casos pueden rebatir componentes de la TdC asumida.
- Si se considera que han tenido la experiencia más típica. Se usa para preguntas descriptivas que indagan sobre la experiencia más común.
- Si se considera que representan casos fuera de los valores esperados en términos de experiencia típica. Estos casos a menudo revisten una importancia crucial para las preguntas sobre la causalidad, porque los casos atípicos pueden rebatir la TdC subyacente. Por ejemplo, una evaluación de un programa de alimentación podría tomar una muestra de madres que abandonaron el programa en una etapa temprana, así como de aquellas que se quedaron la mayor cantidad de tiempo posible en el programa.
- La diversidad de los integrantes de la muestra, de modo tal que el grupo sea inclusivo. Por lo tanto, deberíamos tratar de incluir tanto familias monoparentales como familias nucleares, tanto jóvenes como adultos mayores y así sucesivamente. En las investigaciones de pocos casos o *small-n*, a menudo se toman las muestras de modo tal que sean inclusivas antes que representativas. Esto se debe a que cada caso tiene que añadir significado e información a la investigación.
- Su capacidad para indicar si la teoría del cambio es correcta o no. Estos tipos de casos son nuevamente cruciales para confirmar o rebatir teorías particulares sobre la causalidad. Por ejemplo, si la TdC establecía que la mejora del nivel nutricional dependía no solo de la alimentación, sino también de la higiene y el acceso a atención de salud, podríamos

extraer muestras que cumplieran solo con algunos de estos requisitos para validar la TdC.

## Muestreo por cuotas



### Definición: Muestreo por cuotas

El muestreo por cuotas divide a la población en subcategorías mutuamente excluyentes y luego recoge información para un tamaño o proporción de muestra previamente establecido para cada categoría.

Por ejemplo, si estuviésemos tomando una muestra de hogares de una población que ha recibido asistencia, podríamos establecer una cuota con al menos dos de cada una de las siguientes categorías:

- Hogares con niños pequeños;
- Hogares con ancianos únicamente;
- Hogares encabezados por mujeres;
- Hogares encabezados por hombres;
- Hogares encabezados por niños/niñas;
- Hogares acomodados;
- Hogares pobres;
- Hogares cercanos a la carretera;
- Hogares alejados de la carretera.

Las muestras extraídas bajo cada cuota pueden tomarse intencionalmente o con base en la disponibilidad o incluso en muestreos asistidos. El muestreo por cuotas puede ayudar a garantizar que las muestras incluyan un número suficiente de mujeres, niños y niñas, adultos mayores u otras categorías de interés.



### Consejo útil

Utilicen cuotas si tienen que emplear un muestreo por disponibilidad.

Si no tienen más remedio que emplear un muestreo por disponibilidad, introduzcan cuotas para que la muestra por lo menos sea más inclusiva. Las cuotas pueden usarse en el muestreo por disponibilidad disponiendo, por ejemplo, que cada segunda persona entrevistada sea mujer, cada quinta un hombre joven y así sucesivamente.

### Muestreo con entrevistas asistidas

El uso más común del muestreo con entrevistas asistidas en la EAH es el uso del muestreo de bola de nieve o por referidos, con entrevistas con informantes clave. En el muestreo de bola de nieve se le pide a cada entrevistado que sugiera los nombres de otros posibles entrevistados (Goodman, 1961) que puedan hablar con autoridad sobre el tema. La cadena prosigue hasta que ya no surgen más nombres nuevos. Ver Biernacki y Waldorf (1981) para un buen análisis de esta técnica.

Figura 12.1: Muestreo de bola de nieve



Los informantes referidos en sentido contrario o duplicados no aparecen en el gráfico.

El muestreo de bola de nieve es objeto de diversas críticas (Erickson, 1979) y en ocasiones se aboga por otros enfoques, tales como:

- Muestreo dirigido por los entrevistados (donde se efectúan pagos tanto a la persona que sugiere otros contactos como al otro contacto) (Heckathron, 1997; Wang, 2005);
- Muestreo selectivo (*Targeted sampling*), donde se hace cierta investigación inicial utilizando la técnica de mapeo etnográfico o cartografía etnográfica para identificar a la población muestral (Watters y Biernacki, 1989).

Las críticas al muestreo de bola de nieve y las alternativas se refieren al muestreo de poblaciones ocultas tales como consumidores de drogas por vía intravenosa, antes que al muestreo de directores humanitarios. El muestreo de bola de nieve con frecuencia puede entrañar decenas de entrevistas. Es una buena técnica para subsanar cualquier riesgo no intencional en la selección de los entrevistados iniciales.

La investigación de Guest et al. (2006) sobre el número de entrevistas cualitativas que se requieren para que se produzca una saturación sugiere lo siguiente:

1. El uso del muestreo de bola de nieve para identificar a los entrevistados puede ser ineficiente, ya que después de 6-12 entrevistas con un grupo en particular emergen muy pocos temas nuevos.
2. La pregunta final convencional de una entrevista, “¿Con quién más sugeriría usted que hablásemos en relación con este tema?” probablemente debería complementarse con “Estamos entrevistando a los grupos A, B y C. En su opinión, ¿con qué otros grupos sería conveniente hablar?”



#### Consejo útil

Categoricen a los entrevistados por grupos y apunten a formar grupos de seis a 12 personas cada uno, al mismo tiempo que verifican si siguen surgiendo nuevos temas e informaciones.

## Tamaño de las muestras no aleatorias

El investigador de pocos casos (*small-n*) aún tiene que tomar muchas decisiones en relación con la muestra, no solo en lo que se refiere a la estrategia muestral sino también al tamaño de la muestra. En las muestras *small-n*, la indicación de que el tamaño de la muestra es adecuado suele ser la *saturación de información*. Sin embargo, existe muy poca orientación operativa en términos de cuántas entrevistas se necesitan para llegar a la saturación.

El uso de una herramienta para el cotejo de la evidencia, como el cuadro de evidencia descrito en la Sección 8, Fase inicial, donde la evidencia se registra con base en un esquema de codificación preestablecido, puede ayudar a identificar la saturación. Una herramienta de evidencia de esta naturaleza puede mejorarse aún más a estos efectos, registrando si un argumento que es esgrimido en una entrevista, o por una fuente en particular, es nuevo o simplemente está reforzando los mismos argumentos que ya han sido recogidos con anterioridad. Esto ayudará a identificar la saturación de información con mayor celeridad.

Creswell y Clark (2011: 174) sugieren que para un estudio de caso adecuado se requieren de cuatro a seis casos, y para un estudio basado en entrevistas de 20 a 30 entrevistas. Sin embargo, estas reglas de oro parecen sustentarse en la experiencia práctica de los autores.

Para entrevistas de hogares cualitativas, Few et al. (2014: 58) sugieren un mínimo de 20 entrevistas para una comunidad de 500 hogares y un mínimo de 50 entrevistas para una comunidad de 5,000 hogares, a fin de garantizar que se recoja un rango adecuado de opiniones. Few et al. sugieren asimismo la conveniencia de estratificar dicha muestra, con el fin de garantizar la cobertura de distintos grupos sociales, y para que los hogares puedan ser seleccionados al azar con la finalidad de reducir en la mayor medida posible los sesgos. Cabe señalar que las muestras de este tamaño rara vez pueden ser generalizadas, haciéndolas extensivas a toda la población.

En cuanto a los grupos focales, Morgan (1997) indica que después del tercer grupo emergen pocos temas nuevos. Un estudio sobre la saturación de información en entrevistas con informantes clave (Guest et al., 2006) concluyó que la saturación se alcanzaba a la sexta o duodécima entrevista. En ambos casos, Morgan y Guest están hablando de grupos homogéneos. La mayoría de las intervenciones humanitarias tienen múltiples partes interesadas, cada una de las cuales debería poder hacerse escuchar. Por lo tanto, si ustedes estuvieran estudiando la educación y los jóvenes, por ejemplo, podrían organizar tres grupos focales, uno con hombres jóvenes, otro con mujeres jóvenes y otro con maestros y padres de familia. Asimismo, podrían necesitar entrevistar a una serie de personas, como dirigentes comunitarios locales, representantes de colectivos de mujeres, personal de campo de ONG, directores de ONG y así sucesivamente.

La credibilidad es un tema central de la evaluación. Una evaluación solo puede ser útil si es considerada creíble. Aunque las entrevistas con informantes clave que emplean una muestra de 12 miembros de un grupo dado pueden llegar a la saturación de información, este no es el único factor que debe tomarse en cuenta.

Entrevistas adicionales podrían:

- Reforzar la credibilidad de la evaluación por el hecho de haber consultado las opiniones de un mayor número de personas;
- Reforzar los temas de las primeras 12 entrevistas, proveyendo triangulación en lo que se refiere a dichos temas.

La triangulación se discute en la [Sección 13, Metodologías de campo](#). La triangulación es fundamental en las metodologías cualitativas y mixtas, lo cual significa que las evaluaciones deben trascender el número de entrevistas y casos requeridos para una saturación de información simple.

Un enfoque podría ser el uso de entrevistas grupales, talleres o exámenes retrospectivos como una forma más eficiente de consultar con un mayor número de personas, más allá de las entrevistas con informantes clave. En ese caso los evaluadores deberían esforzarse por alcanzar la saturación de información para cada una de las metodologías utilizadas. Si por diferentes métodos se identifican los mismos temas y problemas, esto constituye un tipo de triangulación para los resultados de la evaluación.

### ¿Qué tiene de malo utilizar un muestreo aleatorio para las metodologías *small-n*?

A primera vista, el uso del muestreo aleatorio debería ayudar a prevenir los sesgos de muestreo, tanto en estudios de pocos casos o *small-n* como en los estudios de un mayor número de casos o *large-n*. Gerring (2007: 87-88) sostiene, sin embargo, que si bien una colección de tamaños de muestra pequeños tendrá en promedio el mismo valor medio que la población, las muestras individuales pueden tener valores medios muy diferentes a los de la población. En consecuencia, las muestras aleatorias de tipo *small-n* pueden arrojar un estimado altamente engañoso de la situación de la población en su conjunto.

En el caso de las metodologías *large-n*, el hecho de que cada nuevo caso añada solamente una pequeña cantidad de información no constituye un problema, ya que la muestra contiene muchos casos. El problema ocurre en los estudios *small-n*, donde se espera que cada caso contribuya significativamente al conocimiento obtenido.

El otro tema es que, dado que la muestra se seleccionó aleatoriamente, puede asumirse equivocadamente que los resultados pueden ser generalizados. No pueden, si el tamaño de la muestra es muy pequeño.



#### Consejo útil

Si tienen que seleccionar pocos casos, no utilicen un muestreo aleatorio o pseudoaleatorio. Seleccionen los casos intencionalmente o mediante el uso de cuotas.

## 12.3 Muestreo aleatorio

El muestreo aleatorio es esencial para las metodologías *large-n* en las que es necesario generalizar a partir de la muestra, haciéndola extensiva a toda la población. Si no se utilizan muestras seleccionadas aleatoriamente, es imposible hacer generalizaciones válidas a partir de una muestra y hacerla extensiva a toda la población, ya que la muestra no es representativa. Un verdadero muestreo aleatorio requiere de un marco muestral y de un listado de toda la población, del cual se pueda extraer la muestra aleatoria. Rara vez se cuenta con un listado completo pero se puede contar con un aproximado, por ejemplo listados de beneficiarios de los cuales se pueda extraer la muestra.

### Muestreo pseudoaleatorio

En muchos casos se pueden utilizar metodologías pseudo-aleatorias en vez de muestreos verdaderamente aleatorios. Esto es común cuando no se cuenta con un marco muestral completo. Las metodologías pseudo-aleatorias incluyen tomar un punto de partida aleatorio (o pseudoaleatorio) y hacer el muestro en un intervalo posterior.



#### **Definición: Muestreo pseudoaleatorio**

Muestreo donde no se cuenta con un marco muestral.

Tradicionalmente, la primera instancia se selecciona aleatoriamente de un punto de partida seleccionado intencionalmente, y las instancias subsiguientes se seleccionan utilizando alguna regla.

Se pueden aplicar diversas reglas, como seleccionar el punto medio de un grupo en un mapa o una fotografía aérea y luego empezar en la dirección establecida por un bolígrafo que se gira sobre la superficie. La construcción de los intervalos se rige por dos reglas: reglas para intervalos fijos (por ejemplo, cuando se selecciona una de cada 10 viviendas – saltándose nueve viviendas) o reglas para intervalos variables, utilizando una hoja o un generador de números aleatorios.



#### **Consejo útil**

Utilicen una aplicación de números aleatorios en su smartphone para generar valores de intervalos aleatorios. Por ejemplo, si desearan entrevistar a una de cada 10 viviendas en promedio, configurarían los valores de los números en un rango de 1 a 17, con la finalidad de generar intervalos aleatorios con un promedio de 9.

Es importante tomar conciencia de que la inferencia estadística se basa en un muestreo verdaderamente aleatorio, y que el uso de técnicas pseudo-aleatorias puede introducir un sesgo desconocido.

## Muestreo por conglomerados y el efecto diseño



### **Definición: Muestreo por conglomerados**

El muestreo por conglomerados es un muestreo en el cual se toma una muestra de un número de ubicaciones, cada una con un conglomerado de un número de casos determinado.

La principal ventaja del muestreo por conglomerados es que puede ser usado sin un marco muestral para personas.

El mecanismo de muestreo por conglomerados común consiste de 30 conglomerados de 30 personas cada uno, pero también pueden usarse otros mecanismos. El muestreo por conglomerados necesita de una muestra global más grande, pero es una técnica efectiva ante la falta de un buen marco muestral o cuando la población se encuentra dispersa en una zona extensa. En este último caso, el muestreo por conglomerados permite un trabajo de campo más rápido y eficiente. El muestreo por conglomerados es ampliamente utilizado para encuestas sobre nutrición e inmunización. El muestreo por conglomerados puede utilizar un muestreo aleatorio o intencional para seleccionar los conglomerados, y un muestreo aleatorio o pseudoaleatorio para la selección dentro del conglomerado.

Utilicen una calculadora de muestras para estimar el tamaño exacto de la muestra que necesitan, y luego añadan un margen de seguridad de al menos 10 % para la falta de respuesta. Si tienen previsto realizar encuestas repetidamente, dejen un espacio para el desgaste entre encuestas.



### **Definición: Efecto de diseño**

El efecto de diseño es el factor por el cual se tiene que modificar el tamaño de una muestra cuando uno se aleja del muestreo aleatorio simple.

Un efecto de diseño de dos (duplicación del tamaño de la muestra) se utiliza tradicionalmente para encuestas de nutrición basadas en muestreos por conglomerados (FIDA, 2014); pero un examen de los efectos de diseño de las muestras por conglomerados (Kaiser et al., 2006) reveló que un efecto de diseño de 1.5 era adecuado para la mayor parte de las encuestas sobre desnutrición aguda. El mismo estudio reveló que las encuestas por conglomerados sobre mortalidad en situaciones de conflicto tenían efectos de diseño de cuatro o más. En su estudio sobre mortalidad durante la crisis alimentaria en Etiopía en 2000, Salama et al. (2001: 564) utilizaron un efecto de diseño de cuatro para la mortalidad.

El efecto de diseño se produce porque, dadas las similitudes entre los miembros de un conglomerado, cada miembro nuevo de un conglomerado proporciona información menos nueva que un nuevo miembro independiente. Por lo tanto, el efecto de diseño es mayor para situaciones en las que los miembros del conglomerado son más proclives a ser similares.

### Muestreo estratificado

La única estrategia de muestreo con un efecto de diseño menor a uno es el muestreo aleatorio estratificado, donde se puede necesitar una muestra marginalmente menor que para el muestreo aleatorio simple.



#### **Definición: Muestreo estratificado**

Muestreo estratificado es una técnica de muestreo donde primero se divide a la población en segmentos mutuamente excluyentes y luego se toma una muestra aleatoria simple de cada segmento.

El muestreo estratificado puede ser de utilidad si se teme que podrían faltar algunas minorías.

### Cómo estimar el tamaño requerido para las muestras aleatorias

Para estimar el tamaño de muestra requerido debemos saber si la muestra se está utilizando para hacer una generalización con respecto a una población o para comparar dos poblaciones distintas.

Debemos saber también si se están examinando las proporciones de una categoría en particular (sufren de desnutrición, no sufren de desnutrición; recibieron un kit con materiales de albergue, no recibieron un kit con materiales de albergue) o los valores medios de una variable (ingresos a nivel de hogares, porcentaje de alimentos de ayuda humanitaria vendidos, consumo de litros de agua diarios).

Es una falacia común afirmar que el tamaño de la muestra debería ser proporcional a la población, pero las muestras relativamente pequeñas pueden ser usadas con poblaciones pequeñas. Daniel sugiere que sería bueno corregir el tamaño real de la muestra cuando ésta represente más del 5 % de la población.

Una vez que la población desciende por debajo de aproximadamente 400 habitantes, el tamaño de muestra requerido sube a más del 50 por ciento de la población (en el peor de los casos). Censar adquiere mayor atractivo a medida que el tamaño de la muestra crece como porcentaje de una población.



### Consejo útil

Utilicen una calculadora en línea para tamaños muestrales para estimar el tamaño de muestra que necesitan. Las calculadoras en línea que presentamos a continuación tienen la ventaja de presentar las fórmulas subyacentes con definiciones de los términos empleados.

- Para estimar la proporción para una población:  
[www.select-statistics.co.uk/sample-size-calculator-proportion](http://www.select-statistics.co.uk/sample-size-calculator-proportion)
- Para estimar la media para una población:  
[www.select-statistics.co.uk/sample-size-calculator-mean](http://www.select-statistics.co.uk/sample-size-calculator-mean)
- Para comparar las proporciones de dos grupos:  
[www.select-statistics.co.uk/sample-size-calculator-two-proportions](http://www.select-statistics.co.uk/sample-size-calculator-two-proportions)
- Para comparar las medias de dos grupos:  
[www.select-statistics.co.uk/sample-size-calculator-two-means](http://www.select-statistics.co.uk/sample-size-calculator-two-means)

Cuando estimamos el valor de algún parámetro de una población con base en una muestra, hacemos una estimación que tiene una probabilidad particular de recaer dentro de un intervalo de confianza dado.



### Definición: Intervalo de confianza

El intervalo de confianza es el rango dentro del cual esperamos que recaiga el valor de una población en su conjunto.

El intervalo de confianza generalmente se expresa como +/- un porcentaje: el más común es +/- 5 %. También se conoce como margen de error. No podemos estar absolutamente seguros, sin embargo, de que el valor de la población recaerá dentro del intervalo de confianza. En un porcentaje reducido de casos puede recaer fuera de este.



### Definición: Nivel de confianza

El nivel de confianza es la probabilidad de que el valor de la población en su totalidad recaiga dentro del intervalo de confianza.

Un nivel de confianza del 95 % significa que hay una probabilidad del 5 % de que el valor de la población recaiga fuera del intervalo de confianza – esto es, una probabilidad del 5 % de obtener un resultado falso positivo.

Cuando estimamos una media poblacional debemos considerar la variabilidad de la población (estimada con base en la variabilidad de la muestra).

**Definición: Desviación estándar**

La desviación estándar o desviación típica es una medida de la variabilidad de un parámetro.<sup>3</sup>

Dos factores adicionales son importantes para comparar diferentes grupos: el tamaño del efecto y el poder estadístico.

**Definición: Tamaño del efecto**

El tamaño del efecto es la diferencia proporcional en la variable de interés entre el grupo tratado y el grupo de control.

Un problema común en la EAH es que los planificadores a menudo sobreestiman el efecto de su intervención. Cuando el nivel del efecto es pequeño se necesitan muestras muy grandes. En sentido inverso, un nivel de efecto grande requiere solo muestras pequeñas.

**Definición: Poder estadístico**

El poder estadístico es la probabilidad de que un resultado negativo no sea un falso negativo (1 menos el riesgo de un resultado falso negativo).

Un poder del 80 % quiere decir un 20 % de probabilidad de no detectar una diferencia significativa, si la hubiera – es decir, un resultado falso negativo. En los ensayos médicos, los resultados falsos negativos suelen ser más aceptables que los resultados falsos positivos; de ahí el tradicional margen de error más amplio. Es cuestionable, sin embargo, que las evaluaciones que deben determinar si las intervenciones han tenido algún impacto puedan usar un poder tan bajo. Un poder del 90 % (con un 10 % de probabilidad de un falso negativo) puede ser más apropiado. La guía de muestreo del Proyecto de Asistencia Técnica para Alimentación y Nutrición sugiere utilizar un poder del 90 % cada vez que los recursos lo permitan (Magnani, 1997).

**Cuadro 12.1:** Factores que influyen en el tamaño de la muestra

Estimación de un resumen estadístico para una población		Comparación de dos poblaciones (ya sea dos grupos o el mismo grupo antes y después)		
	Estimación de una proporción	Estimación de una media	Comparación de dos proporciones	Comparación de dos medias
<b>Nivel de confianza</b>	Típicamente 95 % – un nivel de confianza del 99 % necesita de una muestra mayor en un 70 %, mientras que un nivel de confianza del 90 % necesita de una muestra que sea 30 % menor.			
<b>Intervalo de confianza</b>	Típicamente +/-5 % – para usar +/- 3 % se requiere un tamaño de muestra que sea 2.8 veces el tamaño de la muestra para un intervalo de +/-5 %.			
<b>Proporción base</b>	Una proporción del 50 % necesita de una muestra casi tres veces mayor que una proporción del 10 % o del 90 %.		Una proporción base del 10 % puede requerir un tamaño de muestra que sea 10 veces mayor o más que para una proporción del 50 %.	
<b>Variabilidad</b>		El tamaño de muestra requerido aumenta cuatro veces cuando la desviación estándar se duplica.		El tamaño de muestra requerido aumenta cuatro veces cuando la desviación estándar se duplica.
<b>Tamaño del efecto</b>			El tamaño de muestra requerido aumenta cuatro veces cuando el tamaño del efecto se reduce en un 50 %.	
<b>Poder</b>			Un poder del 90 % requiere de una muestra que sea un 30 % mayor que un poder del 80 % (a un nivel de confianza del 95 %).	

## 12.4 Muestreo para métodos mixtos

Solo las metodologías *large-n* o los censos pueden responder con confianza a preguntas acerca de qué porcentaje de una población obtuvo un beneficio en particular o experimentó un efecto en especial. Solo las metodologías *small-n* pueden ofrecer explicaciones convincentes de por qué esto sucedió. Podría argumentarse que las técnicas *large-n*, acompañadas de un análisis factorial, pueden ofrecer explicaciones, pero solo para aquellos factores explicativos que están explícitamente incluidos en el estudio. La fortaleza de las metodologías cualitativas *small-n* reside en que permiten al evaluador identificar factores y temas que salen a la luz durante la investigación. Esta es una de las razones por las que los enfoques cualitativos son tan útiles en la EAH, en lo que con frecuencia constituyen contextos caóticos.

Las diferentes fortalezas de las metodologías *large-n* y *small-n* para indicar tanto qué pasó como por qué explican por qué las mejores EAH utilizan métodos mixtos.

Los métodos mixtos requieren enfoques de muestreo tanto aleatorios como no aleatorios. Sin embargo, ya sea que utilicen metodologías *large-n* o *small-n*, los evaluadores humanitarios deben explicar su estrategia de muestreo y sus decisiones al respecto. Deben explicar asimismo cualquier potencial sesgo u otras limitaciones inherentes a las decisiones que tomen.

# 13 / Metodologías de campo

La presente sección se centra en las metodologías utilizadas para recabar información, al margen de las metodologías de revisión documental.

Las metodologías de revisión documental se tratan en la [Sección 10, Métodos documentales](#).

Tras una discusión inicial sobre entrevistas en general, analizamos diversas metodologías individuales. Para cada metodología analizamos lo siguiente:

1. El grado en que la metodología se utiliza en la EAH;
2. Cómo se hace;
3. Qué tipo de información se recoge;
4. Qué tipo de muestreo se suele utilizar;
5. Si son caras o de bajo costo;
6. Cuántas se pueden hacer en una EAH pequeña, en la que el trabajo de campo sea de dos semanas.

La sección ofrece asimismo consejos útiles para cada una de las metodologías tratadas.

El análisis de la información recopilada se analiza en la [Sección 16, Análisis](#).

La elaboración de la matriz de evaluación se aborda en la [Sección 8, Fase inicial](#).

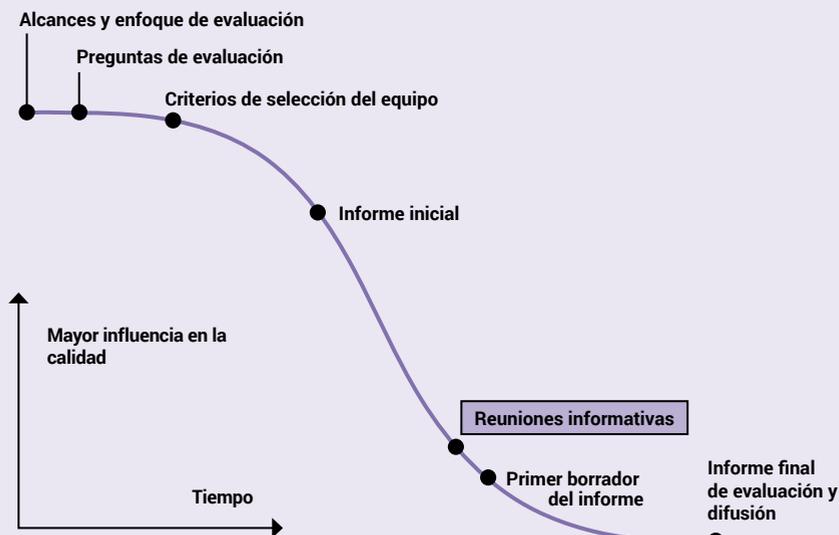


## Tener presente

Las preguntas de evaluación determinan las metodologías utilizadas, pero consideraciones prácticas de índole presupuestal, temporal, ética y logística pueden interferir con el uso de las metodologías más apropiadas para una pregunta en particular.

Las evaluaciones pueden aplicar una variedad de metodologías para medir lo que desean. Por ejemplo, la evaluación de IRC de su proyecto de reconstrucción en el condado de Lofa utilizó juegos conductuales así como encuestas para determinar si las comunidades beneficiadas por el proyecto se comportaban de manera diferente que las comunidades de control: “Buscamos mediciones directas de los efectos conductuales del programa de reconstrucción dirigida por la comunidad (RDC), tal como lo revelaron las conductas desplegadas en los ‘juegos de bienes públicos.’” (Fearon et al., 2008: 25).

## Puesto de control de calidad: Reuniones informativas



Las reuniones sobre el terreno para informar sobre los hallazgos pueden contribuir a identificar importantes vacíos e influir en el análisis de la información recogida.

## 13.1 La entrevista como técnica

La entrevista es la base de muchas metodologías de investigación y puede ser dividida en tres categorías:

1. **Entrevistas estructuradas.** Utilizadas en las encuestas cuantitativas. Los entrevistadores se ciñen a un protocolo estricto para hacer las preguntas.
2. **Entrevistas semiestructuradas.** Ampliamente utilizadas en la EAH. El entrevistador realiza preguntas al entrevistado para extraer información y significados.
3. **Entrevistas no estructuradas.** La entrevista consiste en una conversación abierta, aunque puede tener una temática central, por ejemplo una historia de vida. Poco usadas en la EAH.

Una entrevista es un diálogo, y no solo una transferencia de información. Esto es tan cierto para las entrevistas formales realizadas en el marco de encuestas como para las entrevistas cualitativas con informantes clave (Suchman y Jordan, 1990).

Las investigaciones sobre entrevistas formales para encuestas demuestran que las respuestas varían, no solo dependiendo de cómo se formule la pregunta sino también de acuerdo al género, origen étnico, edad, nivel socioeconómico aparente y contextura física del entrevistador (Dohrenwend et al., 1968; Eisinga et al., 2011; Flores-Macias y Lawson, 2008; Kane y Macaulay, 1993; Riessman, 1977; Suchman y Jordan, 1990; Webster, 1996).



### Tener presente

En algunos contextos, entrevistar a las personas puede representar una amenaza para su integridad. No realicen entrevistas en dichas circunstancias, salvo que los entrevistados estén plenamente conscientes de los riesgos y los acepten. Aunque la entrevista los puede poner en riesgo, también tienen derecho a hacerse escuchar (ver la [Sección 14, Involucrando a la población afectada en su evaluación](#)).

## Entrevistas estructuradas

Las entrevistas estructuradas normalmente se llevan a cabo en el marco de encuestas cuantitativas. Para reducir los sesgos en la mayor medida posible, y por extensión los errores en las estimaciones, las entrevistas se llevan a cabo de manera estandarizada. Los entrevistadores no indagan más allá de las preguntas establecidas.

A los entrevistadores a menudo se les conoce como encuestadores y pueden ser facilitados por una empresa de investigación de mercados que esté realizando la encuesta, o bien pueden ser estudiantes universitarios. Los encuestadores no necesitan el mismo nivel de habilidades y comprensión del contexto que los entrevistadores de un estudio cualitativo y por lo tanto tienden a ser menos caros de contratar.

Los encuestadores deben hacer las preguntas en una secuencia establecida y abstenerse de explicar la pregunta, para evitar sesgar las respuestas ya que otros encuestados podrían haber dado una respuesta diferente si hubieran recibido la misma explicación. Para subrayar su formalidad, el cuestionario utilizado es conocido como el instrumento de la encuesta.



### Definición: El instrumento de la encuesta

El instrumento de la encuesta es el cuestionario que utiliza el entrevistador en las entrevistas formales de las encuestas.

El cuestionario de la encuesta puede tener un gran número de preguntas, organizadas en grupos para examinar diferentes hipótesis o conceptos tales como liderazgo o capacidad. El cuestionario de la encuesta a los maestros para la revisión del programa de educación en Palestina del NRC (Shah, 2014) constaba de 91 preguntas para medir 10 hipótesis subyacentes. Un elemento estándar del cuestionario de la encuesta es comprobar que las preguntas se articulan bien con las siguientes en lo que se refiere a una hipótesis en particular (la confiabilidad interna).

Las preguntas se extraen del cuestionario. Es fundamental validar el cuestionario previamente con el objeto de identificar las preguntas que generan preguntas o respuestas inciertas y reformularlas (ver el [ejemplo de buena práctica en la página 251](#) para la experiencia de World Vision Internacional [WVI] sobre traducción de ida y vuelta).

Si bien pueden incluir algunas preguntas abiertas, las entrevistas estructuradas presentan dos problemas:

- La respuesta tiene que ser codificada de alguna manera para poder ser analizada, lo cual puede ser caro y requiere tiempo.
- La respuesta a la pregunta abierta puede acarrear otras preguntas que el encuestador no está autorizado o capacitado para responder.

### Entrevistas semiestructuradas

Las entrevistas semiestructuradas se encuentran en la base de una gama de metodologías ampliamente utilizadas en la EAH. En las entrevistas semiestructuradas el entrevistador utiliza una guía para entrevistas o temas y formula las preguntas en torno a ella. En este enfoque, el entrevistador profundiza en las respuestas previas e indaga más allá con la intención de extraer información y significados.

Las preguntas de sondeo<sup>4</sup> son preguntas que se plantean con la intención de:

- **Aclarar las respuestas** – cuando un entrevistador no está seguro de qué es lo que quiso decir el entrevistado;
- **Obtener mayores detalles** – el uso más común de las preguntas de sondeo en las entrevistas del sector de EAH;
- **Recabar el análisis de los entrevistados** – pueden ser muy útiles para entender por qué las personas actuaron de cierta manera;
- **Preguntar sobre variaciones y casos contrafactuales** – útiles para sacar a relucir la lógica detrás de las decisiones adoptadas.

Las preguntas de sondeo que se utilicen dependerán de las preguntas de evaluación y del alcance de la evaluación.

### Preguntas de sondeo aclaratorias

Estas pueden incluir:

- ¿Le he entendido bien o dijo usted que...? Esta es a menudo la mejor pregunta aclaratoria, ya que se cuida de sugerir que el entrevistado pudiera haberse expresado con falta de claridad en su respuesta anterior.
- Cuando usted dice... ¿qué quiere decir exactamente?
- Usted utilizó el término... ¿Estoy en lo correcto si digo que se está refiriendo usted a...?

### Preguntas de sondeo para pedir más detalles

Estas pueden incluir:

- ¿Me puede dar un ejemplo?
- ¿Cómo lidió con eso?
- ¿Qué respuesta obtuvo?
- ¿Me puede decir algo más acerca de su experiencia?

### Preguntas de sondeo analíticas

Estas pueden incluir:

- ¿Cómo caracterizaría usted lo que sucedió?
- ¿De ello qué sobresale en su recuerdo?
- ¿Qué era lo importante?
- ¿Cuáles esperaba usted que fueran los resultados?
- ¿Cómo se sintió usted con respecto a ello?

### Variaciones y preguntas de sondeo contrafactuales

Estas pueden incluir:

- ¿Su enfoque ha variado con el tiempo? ¿En qué sentido ha variado? ¿Qué propició dicho cambio?
- ¿Lidiaría con esta situación de la misma manera la próxima vez?
- Me doy cuenta de que ustedes están haciendo... pero la organización Y está haciendo... con relación al mismo tema. ¿Qué ventajas tiene su enfoque frente al de ellos y viceversa?
- Algunas de las personas con las que he hablado dijeron... ¿Qué opina usted al respecto?

### Entrevistas no estructuradas

Estas entrevistas pueden abarcar desde entrevistas no estructuradas de ninguna manera como las que se pueden utilizar en la investigación etnográfica hasta entrevistas temáticas, donde el entrevistador introduce un tema – por ejemplo, la historia de vida del entrevistado – y deja que el entrevistado se exprese al respecto.

Las entrevistas no estructuradas no han sido utilizadas muy frecuentemente en la EAH, pero algunas técnicas de entrevista, como las discusiones de grupos focales, incluyen un componente de la entrevista no estructurada. Un ejemplo del uso de la entrevista temática no estructurada fue la evaluación del Centro de Asistencia a Desastres (DAC, por sus siglas en inglés) de la respuesta a las inundaciones de Mozambique en 2000. Aquí, un evaluador efectuó largas entrevistas a miembros de la población afectada, para formarse una idea de su experiencia con la respuesta de emergencia (Cosgrave et al., 2001).

## Los sesgos en la entrevista

Sin duda, todo entrevistador y todo entrevistado tiene una agenda propia. Algunos entrevistados de organizaciones pueden desear que su trabajo se conozca más ampliamente; otros pueden querer ocultar defectos. La población afectada puede desear que continúe la asistencia, o puede preferir que se le ponga fin. El evaluador debe tener presente en todo momento la posibilidad de que dichos sesgos ocurran, y que las entrevistas son procesos construidos que se mueven entre los intereses y valores diferentes del entrevistador y el entrevistado. El Capítulo 16, Análisis, ofrece información sobre cómo manejar los sesgos en la evidencia de la entrevista.

Como sucede con todas las metodologías cualitativas o mixtas, la triangulación es la técnica más eficaz para controlar los sesgos. La triangulación es más poderosa cuando la selección intencional de los entrevistados es lo más inclusiva posible. La triangulación para manejar los sesgos en las entrevistas, y las controversias en torno a la información, es abordada de manera más detallada en la [Sección 13, Metodologías de campo](#).



### Consejo útil

Al llevar a cabo las entrevistas con la población afectada, el horario en que se realizan las entrevistas tiene una gran influencia sobre quiénes estarán disponibles para ser entrevistados. Los evaluadores deben ser conscientes de las personas que están excluyendo por entrevistar a ciertas horas del día.

Ver la [Sección 11: Diseños de evaluación para responder las preguntas de evaluación](#), para sesgos en general.

## 13.2 Intérpretes

Idealmente, los evaluadores deberían ser capaces de hablar el idioma de las personas que están entrevistando. Si no es así, necesitarán trabajar con intérpretes para algunos tipos de entrevistas. La interpretación siempre ocasiona problemas en la evaluación. No está exenta de errores, y Temple et al. (2006) sostienen que entrevistar a través de un intérprete es más parecido a analizar información secundaria que a realizar una investigación primaria.

Aunque pueden usarse intérpretes para las entrevistas semiestructuradas y entrevistas grupales, es casi imposible usar intérpretes para entrevistas estructuradas o discusiones de grupos focales eficaces. Las entrevistas realizadas en el marco de encuestas ya de por sí toman un tiempo considerable, y añadir un tiempo adicional para la interpretación las extiende aún más. En los grupos focales, la mecánica de la interpretación interfiere con el concepto de intercambios que fluyan libremente.

En el caso de los grupos focales, Steward et al. (2007: 112) afirman que: Los moderadores deberían hablar el idioma materno en que se desenvuelve el grupo. Hasta la persona más diestra en una segunda lengua tendrá cierta dificultad para formular las preguntas y seguir las respuestas. Los intérpretes generan problemas adicionales y deben ser evitados siempre que sea posible. A menudo sucede que las percepciones y opiniones del intérprete adornan la traducción.

Tener que depender de un intérprete incrementa el tiempo requerido para lidiar con un conjunto de preguntas. En ocasiones esto puede ser útil si se trata de entrevistas cualitativas, ya que concede más tiempo al evaluador para reflexionar sobre las respuestas a las preguntas previas al mismo tiempo que dilucida las preguntas siguientes. Los evaluadores deben hacer todos los esfuerzos posibles para contratar intérpretes calificados. Un intérprete versado e informado también puede añadir valor, al resaltar aspectos contextuales de las discusiones después de la entrevista que podrían haber pasado desapercibidos para el evaluador.



#### Consejo útil

Consignan a los mejores intérpretes que puedan permitirse. Los intérpretes profesionales valen sus honorarios, ya que son menos propensos a distorsionar la pregunta o las respuestas. De modo que incluyan este gasto en su presupuesto.

Recuerden que reclutar intérpretes toma tiempo. Los intérpretes deben ser incluidos en la tramitación de los viajes y el alojamiento.

En una evaluación en Afganistán, se preguntó a los ancianos de la comunidad qué impacto habían tenido los talibanes en el acceso de la población a los servicios del Estado. El miembro del personal de la organización que fungía de intérprete le dijo al equipo que los ancianos se habían quejado de que las acciones de los talibanes habían reducido su acceso a los servicios. Más adelante, el asistente del equipo de evaluación le dijo al equipo que lo que los notables realmente habían dicho era que ellos *eran* los talibanes y que el gobierno rehusaba prestarles servicios si no entregaban sus armas.



#### Consejo útil

Eviten utilizar al personal de la organización para el trabajo de interpretación siempre que sea posible.

Aunque está lejos de ser lo ideal, no siempre se puede evitar utilizar al personal de la organización para la labor de interpretación, ya sea porque no hay intérpretes profesionales disponibles para los idiomas requeridos o debido a las restricciones de seguridad o acceso.

En la evaluación de la respuesta de CARE a la crisis humanitaria de 2011-2012 en el Sahel (Gubbels y Bousquet, 2013: 18), el equipo dejó constancia de que en las discusiones de grupos focales se vio obligado a recurrir a uno de los animadores de CARE en Chad para que fungiera de intérprete. Esto representaba una potencial fuente de sesgo. Sin embargo, al utilizar preguntas de sondeo cada vez que fuera necesario, y evaluar el patrón general de las respuestas de seis aldeas, el equipo considera que se evitó el problema del sesgo positivo.



#### Consejo útil

Constaten por distintos medios la interpretación realizada por personas que pudieran tener intereses creados. Si no conocen el idioma y se ven obligados a recurrir al personal de la organización para interpretar, es recomendable que haya un asistente presente que entienda el idioma y pueda alertar al equipo sobre errores graves posteriormente. Otra opción es grabar abiertamente los intercambios para su análisis posterior, lo cual servirá de incentivo para interpretar correctamente.

## 13.3 Métodos de entrevista cualitativa

### Entrevistas con informantes clave

Las entrevistas con informantes clave constituyen la piedra angular de la EAH.

#### Características principales:

- Utilizan una técnica semiestructurada y una guía para entrevistas.
- Suelen durar entre 45 y 60 minutos.
- Arrojan información cualitativa, que es la mejor para temas de significado y percepción.
- Son efectivas para responder preguntas acerca del cómo y el por qué sucedieron las cosas.
- No son efectivas para recopilar información cuantitativa.
- Requieren entrevistadores experimentados, que puedan hacer preguntas de sondeo para sacar a relucir los significados.
- Normalmente se registran por escrito, pero también pueden registrarse en un dispositivo digital.
- El muestreo suele ser intencional. Se selecciona a los entrevistados considerados más idóneos para responder las preguntas de evaluación.
- Se trata de una metodología rápida y relativamente económica.
- En dos semanas de trabajo de campo se podrían realizar entre 25 y 50 entrevistas con informantes clave.
- El número de entrevistas promedio no suele exceder las cuatro o cinco entrevistas diarias.

- La saturación teórica (las entrevistas subsiguientes no arrojan información nueva) para un tipo específico de entrevistado ocurre entre la sexta y la duodécima entrevista de un tipo de entrevista en particular (Guest et al., 2011) pero es posible que se requieran más entrevistas para reforzar la credibilidad.



### Tener presente

Los aspectos éticos más importantes que hay que tener en cuenta son que las entrevistas no deben poner en riesgo al entrevistado y que se debe respetar la confidencialidad.



### Consejos útiles

Las entrevistas se pueden acelerar ligeramente informando a los entrevistados por anticipado cuáles son los objetivos y las reglas básicas que rigen la entrevista.

Las entrevistas con informantes clave suelen tener lugar en un espacio privado, para impedir que otras personas escuchen o se sumen. También se pueden utilizar espacios públicos que tengan normas de privacidad – por ejemplo, un café.

El evaluador a menudo es percibido como un representante de los donantes. En muchas culturas es de mala educación expresarse críticamente frente a alguien que es percibido como representante de la organización que dio la ayuda. En dichas culturas puede ser más útil preguntar al entrevistado qué opinan sus vecinos de la asistencia provista.



## En profundidad: Guía de entrevistas

La guía de entrevistas establece la tendencia general de las preguntas, empezando por recordar al entrevistador que debe informar al entrevistado cuál es el propósito de la evaluación y cuáles son las reglas básicas de la entrevista (tradicionalmente, los comentarios vertidos en una entrevista no son atribuibles, ni directa ni indirectamente).

Es conveniente empezar por preguntar al entrevistado qué cargo ocupa – las personas casi siempre están felices de hablar sobre sí mismas, incluso cuando son tímidas. No todas las preguntas de la guía serán relevantes para todos los entrevistados, y el entrevistador no debe hacer preguntas irrelevantes.

Las preguntas deben hacerse en la secuencia adecuada para establecer una buena relación – evitando las preguntas más polémicas hasta más adelante en la entrevista.

La guía de la entrevista debe cubrir entre 20 y 25 preguntas como máximo. Si hubiera más preguntas, será necesario priorizar o elaborar guías separadas para diferentes tipos de entrevistados.

Las preguntas finales deben incluir una pregunta acerca de lo que han aprendido o han visto reforzado por la respuesta. Preguntar al entrevistado si le sorprendió que no se tocara algún tema puede ser útil para entender cuál es su idea de la intención de la evaluación y en ocasiones puede arrojar pistas nuevas.

En las preguntas finales se debe pedir asimismo al entrevistado que sugiera otras personas y grupos para entrevistar. Pregunten si el entrevistado estaría disponible para responder a preguntas adicionales más adelante en el proceso si fuera necesario.



### Consejo útil

Eviten las preguntas que suscitan una respuesta “correcta”. Por ejemplo, si desean saber si el equipo ejecutor tomó en cuenta el enfoque de género en la intervención, no pregunten: “¿Tomaron en cuenta el enfoque de género?” ya que la respuesta será siempre que sí. En vez de ello, pregunten: “¿Se beneficiaron en mayor grado las mujeres o los hombres de la intervención y, de ser así, por qué? ¿Qué diferencias hubo entre los hombres y las mujeres en lo que respecta a los efectos?”.

## Entrevistas de hogares

Las entrevistas de hogares se utilizan habitualmente en el marco de encuestas cuantitativas, pero también pueden usarse para recoger información cualitativa.

### Características principales:

- Pueden emplear un instrumento formal para encuestas o una guía para entrevistas, o bien una combinación de ambos.
- Pueden tomar entre 20 y 60 minutos, dependiendo del grado de complejidad de la entrevista.
- Pueden recolectar información cualitativa y cuantitativa, dependiendo del tipo de entrevista.
- Son efectivas para responder a preguntas acerca del efecto de la asistencia en los hogares.
- Requieren entrevistadores que tengan un buen conocimiento del contexto y que mediante la observación puedan detectar indicios para triangular la información recibida.
- Normalmente registradas en los formularios de encuestas, pero pueden registrarse también en un dispositivo digital.
- La ubicación de las entrevistas se selecciona intencionalmente, pero los hogares pueden ser seleccionados ya sea aleatoriamente (para estudios *large-n*) o intencionalmente para estudios *small-n*.
- La metodología es moderadamente cara, salvo que sea parte de un sistema de seguimiento regular.
- Normalmente se llevan a cabo como un ejercicio separado si la intención es que sean representativas.
- El número de entrevistas diarias depende de la complejidad y del tiempo de viaje.

Cuando se utilizan para recabar información cualitativa, normalmente inciden en unos cuantos temas solamente y son mucho más cortas que una entrevista en regla con informantes clave. Su ventaja más importante sobre las entrevistas grupales es que el hogar constituye un espacio privado que hace posible abordar temas más sensibles.



#### Consejo útil

Las entrevistas de hogares pueden arrojar información efectiva para casos ilustrativos que inciden en temas puntuales. Dichos casos ilustrativos sirven para que los lectores tomen conciencia de la dimensión humana y agilizan la lectura del informe de evaluación.

## Entrevistas grupales

### Características principales:

- Utilizan una técnica semiestructurada así como una guía para entrevistas incompletas, ya que se puede discutir un máximo de cinco a 10 preguntas.
- Normalmente duran entre 30 y 45 minutos, pero pueden durar más si los entrevistados se muestran interesados.
- Recogen información cualitativa, que es la mejor para temas de significado y comprensión.
- Son efectivas para responder preguntas acerca de las percepciones de la comunidad con respecto a la ayuda.
- Son poco eficaces para recolectar información cuantitativa.
- Se necesitan entrevistadores experimentados que puedan relacionarse con la comunidad y formulen preguntas de sondeo para extraer significados.
- Suelen registrarse en forma escrita, pero pueden ser registradas en un dispositivo digital.
- La ubicación de la entrevista se selecciona intencionalmente y los entrevistados iniciales también pueden ser seleccionados intencionalmente; sin embargo, en vista de que suelen realizarse en un espacio público, otras personas pueden sumarse.
- La metodología es rápida y relativamente económica.
- En dos semanas de trabajo de campo podrían efectuarse entre cinco y 20 entrevistas grupales.
- Típicamente no más de dos o tres entrevistas por día.
- No existen estimaciones estandarizadas de la saturación teórica en dichas entrevistas, pero se puede esperar que esta ocurra después de dos o tres entrevistas para grupos focales y después de seis a 12 entrevistas para entrevistas con informantes clave.
- Tienden a ser dominadas por algunas personas, en especial en sociedades jerarquizadas.



#### Tener presente

Las entrevistas grupales pueden provocar objeciones éticas de envergadura debido a que se llevan a cabo en lugares públicos y hablar en público puede amenazar la integridad de las personas entrevistadas.



### Consejos útiles

- Manejen al grupo. Si se reúnen en un lugar al aire libre, no permanezcan de pie: siéntense en el suelo de ser posible. Esto obligará al grupo a sentarse también, evitando con ello posibles empujones.
- Averigüen cuáles son las normas sociales locales y síganlas en la medida de lo posible.
- Encaucen las preguntas, alejándolas de los participantes más locuaces. Pregunten cuántos están de acuerdo con un punto de vista en particular.

## Reuniones formales con estructuras comunitarias

Una variante de la reunión grupal que es muy común en algunas regiones de Asia es una asamblea organizada por las estructuras de la comunidad. Estas pueden ser bastante formales, con una mesa de autoridades y una persona presidiendo la reunión. Estas reuniones tienden a ser dominadas por los dirigentes comunales, lo cual puede dificultar la recogida de información de la comunidad más amplia.

El mejor enfoque consiste en hacer preguntas que en su opinión tendrán respuestas diferentes de personas diferentes del grupo. Preguntar, por ejemplo, quiénes fueron los más afectados, si los hombres o las mujeres se beneficiaron más, qué tipo de asistencia fue de mayor utilidad durante la primera semana, debería sacar a relucir algunas diferencias de opinión. Empezar a recibir respuestas espontáneas puede requerir algo de esfuerzo pero, una vez que echan a rodar, dichas reuniones pueden suministrar información valiosa.

11 / Diseños de evaluación para responder las preguntas de evaluación 192 - 213

12 / Muestreo 214- 230

13 / Metodologías de campo 231 - 267

14 / Involucrando a la población afectada en su evaluación 268- 282

15 / Acceso restringido 283 - 292

16 / Análisis 293 - 321



### Ejemplo de buena práctica: Consultando con la población afectada

En una evaluación de ACNUR sobre su programa de transversalización del enfoque de edad, género y diversidad (Thomas y Beck, 2010), el estudio nacional de Colombia empleó un enfoque altamente participativo. Se seleccionó a cuatro comunidades de diferentes regiones del país para que participaran en la evaluación, consistente de las siguientes actividades (Mendoza y Thomas, 2009):

- Un taller con miembros de la comunidad para elaborar una línea de tiempo de los acontecimientos que antecedieron a la evaluación participativa de ACNUR (una herramienta central del proceso de transversalización) y el plan de acción resultante – para estimar el nivel de conocimiento del plan de acción y evaluar si la evaluación participativa había generado cambios en la comunidad.
- Reuniones con subgrupos de mujeres, hombres, adolescentes de ambos sexos, niños y niñas, adultos mayores y personas con discapacidad – para establecer si diferentes grupos albergaban percepciones diferentes de los resultados de la evaluación participativa y del plan de acción.

## Discusiones de grupos focales

Si bien el término “grupo focal” se aplica muy ampliamente a las entrevistas grupales, es mejor utilizarlo solamente para entrevistas estructuradas.

### Características principales:

- La duración suele ser de dos horas.
- Normalmente de seis a ocho participantes.
- Basadas en una guía temática que establece entre tres y cinco temas de discusión.
- Máxima eficacia con el personal de la organización contraparte, pero también pueden usarse con la población afectada.
- Generan información cualitativa, eficaz para explorar los puntos de vista de grupos específicos (padres y madres solteras, jóvenes, etc.).
- Necesitan un espacio cómodo, controlado.
- Necesitan un facilitador y alguien que tome apuntes.
- Suelen ser registradas en video o grabadas (ver la [Sección 14, Involucrando a la población afectada en su evaluación](#)).
- Las muestras pueden tomarse aleatoriamente de los grupos para formar el grupo focal. Todos los miembros del grupo deben ser similares entre sí, sin diferencias de nivel socioeconómico u otras.
- Morgan (1997) advierte que pocos temas nuevos emergen después del tercer grupo focal.
- Costo intermedio.

- Típicamente menos de 10 discusiones de grupos focales en una evaluación de dos semanas de duración.
- En algunos países se puede remunerar a los participantes de un grupo focal por el tiempo invertido. Este no suele ser el caso en la EAH, pero al menos se les debe ofrecer refrescos por lo que constituye un compromiso de tiempo relativamente largo.

Las referencias habituales para grupos focales son: Krueger y Casey (2009) y Steward et al. (2007).



**Consejo útil**

Interactúen con el grupo antes de iniciar la discusión, para identificar a los más habladores. Tomen un refrigerio juntos antes de empezar la discusión, para identificar a quienes son más propensos a dominar la discusión. Ubíquenos junto al facilitador para mantener menos contacto visual con ellos. Ubiquen a los más callados frente al facilitador, para hacer contacto visual con ellos con mayor frecuencia.

Al analizar la copia registrada del grupo focal, estarán examinando factores tales como palabras, contexto, cambios de opinión, frecuencia de ciertos comentarios, volumen de debate, intensidad y especificidad tanto como el contenido.



**Ejemplo de buena práctica: El uso de discusiones de grupos focales por World Vision International**

Para las discusiones de grupos focales (DGF) normalmente contamos con un jefe de equipo en cada equipo, aunque ocasionalmente no nos es posible, en cuyo caso recurrimos a la estructura regional (podrían ser dos o tres equipos de DGF en una región, con un líder de equipo apoyándolos a todos, por ejemplo). Pero nuestros equipos de DGF están integrados por cuatro personas como mínimo, y seis como máximo: un jefe del equipo, un traductor, dos personas que toman apuntes, un observador y un facilitador. El observador y el facilitador son de sexos opuestos y se intercambian el rol de facilitadores dependiendo del género de la DGF; el observador, el jefe del equipo y el traductor pueden asumir al mismo tiempo la tarea de tomar apuntes si la discusión resulta muy animada.

Fuente: Kathy Duryee, WVI, comunicación personal, 2014

## 13.4 Métodos de encuesta

### Uso de encuestas

Las encuestas, especialmente las encuestas formales frente a frente, son consideradas una fuente de información muy autorizada. Esto se debe en parte a que muchos lectores tienen un conocimiento limitado de estadística. Tomemos el siguiente ejemplo:



#### **En profundidad: Sondeo de opinión política reporta una caída de 2 % en el apoyo a un partido... ¿Qué significa esto?**

Muchos sondeos de opinión política utilizan un nivel de confianza del 95 % y un margen de error de +/- 3 %. Este sondeo revela que el apoyo al partido de gobierno ha caído, del 35 % en la última encuesta a 33 % en la actualidad.

#### **Supuesto común:**

El 33 % de la población apoya actualmente al partido de gobierno.

- **Realidad:**

El 33 % de la muestra apoyaba al partido de gobierno.

#### **Supuesto común:**

Existe una probabilidad del 95 % de que el 33 % de la población apoye al partido de gobierno.

- **Realidad:**

Existe una probabilidad del 95 % de que entre el 30 % y el 36 % (33 % +/- 3 %) de la población apoye al partido de gobierno. Sigue habiendo una probabilidad del 5 % de que el apoyo sea menor al 30 % o mayor al 36 %.

#### **Supuesto común:**

El apoyo al partido de gobierno ha bajado.

- **Realidad:**

En la encuesta pasada, el nivel de apoyo de la población al partido de gobierno tenía un 95 % de probabilidades de ubicarse entre 32 % y 38 % (35 % +/- 3 %), mientras que en la actual se sitúa entre el 30 % y el 36 %. Incluso es concebible que el apoyo popular haya subido, del 32 % al 36 %. Sin embargo, no todos los resultados son igualmente probables, dado que las medias muestrales normalmente se distribuyen alrededor de la media poblacional. El resultado es que es más probable que la media poblacional esté cerca de la media muestral que de los límites. Es por eso que las encuestas de opinión rara vez revelan grandes cambios.

Es fundamental saber algo de estadística para poder entender los informes de las encuestas. Mientras que los equipos pueden informar de márgenes de error superpuestos en las encuestas que miden situaciones antes y después, se enfocan en los cambios en el valor medio de la variable de interés, en vez de hacer ver que es posible que no haya habido ningún cambio.

## Encuestas de nutrición

Las encuestas de nutrición son ampliamente utilizadas para monitorear los programas de nutrición. Las evaluaciones tienden a utilizar información de encuestas existente, antes de llevar a cabo su propia encuesta, como en el caso de la evaluación de Ruanda de la asistencia alimentaria para los refugiados de larga data (Sutter et al., 2012). Dichas encuestas son bastante técnicas y no las analizamos en mayor profundidad aquí.

## Encuestas formales frente a frente

Este tipo de encuestas puede utilizar personas individuales como punto de muestreo, pero en la EAH es más común utilizar hogares como punto de muestreo. La unidad de análisis puede ser ya sea el hogar o un miembro del hogar, que se selecciona utilizando una técnica pseudo-aleatoria como el cumpleaños más reciente, el método de Kish<sup>5</sup> o el enfoque de muestreo por cuotas (alternando hombres y mujeres para garantizar un balance de género). En ocasiones, sin embargo, hablar con alguien que no sea el informante inicial puede representar culturalmente un problema (McBurney, 1988).

### Características principales:

- Utilizan un instrumento de muestreo al que los entrevistadores se ciñen rigurosamente.
- La duración de cada entrevista depende de la extensión y la complejidad del cuestionario de la encuesta.
- Recogen información cuantitativa – efectiva para responder las preguntas que indagan sobre el “cuántos”. También pueden hacer preguntas cerradas de corte cualitativo, ya que las preguntas abiertas son sumamente demandantes a nivel de codificación y análisis.
- Los encuestadores tradicionalmente reciben solo un par de días de capacitación.
- La información se registra en un formulario de la encuesta o en un dispositivo digital. El registro electrónico es muy superior.
- Los entrevistados se seleccionan aleatoriamente para que los resultados de la encuesta puedan ser generalizados, haciéndolos extensivos a toda la población. Cada desviación de un muestreo verdaderamente aleatorio (por ejemplo, un muestreo pseudoaleatorio o por conglomerados) aumenta el tamaño de muestra requerido para arrojar resultados estadísticamente válidos.
- Habitualmente se requieren varios cientos de entrevistas, pero pueden llegar a ser varios miles.
- Son costosas debido a la necesidad de verificar y validar la encuesta y al gran número de entrevistas requerido. Su uso en la EAH tiende a limitarse a evaluaciones a gran escala que cuentan con un presupuesto elevado.

Las encuestas frente a frente pueden ser un poderoso instrumento para determinar la extensión de algunos temas, entre ellos la escala del efecto de una intervención.

Las encuestas con la población afectada son muy eficaces para revelar las percepciones frente a la ayuda, los límites de la cobertura, la conveniencia de las fechas, etc. Si pueden costearlo y no han convocado ya a la población afectada por otros medios, vale la pena realizar una encuesta de la población afectada.

El tamaño de muestra requerido varía, dependiendo de una serie de factores, entre ellos si desean hablar acerca de la prevalencia en un grupo solamente y a través del tiempo o si desean comparar dos grupos diferentes. Ver la Sección 11: Diseños de evaluación para responder las preguntas de evaluación, para consejos sobre el muestreo.

Fowler (2009) ofrece una buena introducción sucinta al tema de las encuestas.

Las encuestas formales están cobrando mayor popularidad en la EAH, especialmente en el caso de evaluaciones grandes con un buen respaldo financiero.

## Ejemplos de encuestas en la EAH

Los casos que presentamos a continuación son ejemplos del uso de muestras formales en la EAH:

- Las encuestas de la población afectada realizadas tras el tsunami de 2004 en el Océano Índico, el terremoto de 2006 en Pakistán y el terremoto de 2007 en Yogyakarta (Bliss y Campbell, 2007a, 2007b; Bliss et al., 2006; Fritz Institute, 2005a, 2005b, 2005c).
- El estudio para vincular la ayuda de emergencia, la rehabilitación y el desarrollo realizado por la Coalición para la Evaluación del Tsunami (Christoplos, 2006) con 1,227 encuestados en Indonesia y 915 en Sri Lanka.
- El estudio de seguimiento de la vinculación entre la ayuda de emergencia, la rehabilitación y el desarrollo (Brusset et al., 2009) realizado con 1,178 encuestados en Indonesia y 965 encuestados en Sri Lanka.
- La evaluación interinstitucional del apoyo a los refugiados en situación de desplazamiento prolongado presentó encuestas en cada uno de los países escogidos como estudios de caso, entre ellos Etiopía con un tamaño de muestra de 1,081 encuestas (Sutter et al., 2011), Ruanda con una muestra de 1,200 encuestas (Sutter et al., 2012) y Bangladesh con una muestra de 1,069 encuestas (Nielsen et al., 2012).

Una vez recolectada, la información de la encuesta debe ser digitalizada y depurada. Este proceso puede ser largo y costoso, especialmente si la encuesta contiene preguntas abiertas.



### Consejo útil

Utilicen dispositivos electrónicos para recolectar las respuestas de las encuestas. Esto eliminará el costo de digitalizar la información y gran parte del costo de depurarla, ya que el formulario de la encuesta puede ser configurado para rechazar combinaciones inválidas en las respuestas (por ejemplo, un informante de 16 años que tiene 12 hijos), y la información recolectada puede simplemente subirse a la base de datos principal sin necesidad de digitalización adicional. Esto puede acelerar el proceso de la encuesta.

Las herramientas digitales pueden emplearse en las encuestas realizadas para diagnósticos y en las encuestas para evaluaciones. El Mecanismo Multiclúster de Diagnóstico Rápido en Pakistán utilizó asistentes digitales personales (PDA, por sus siglas en inglés) para recolectar información para su diagnóstico (Equipo de McRAM, 2009). Shirima et al. ofrecen un ejemplo del uso de PDA en una encuesta de 21,000 hogares en Tanzania e incluyen algunos detalles sobre los aspectos prácticos involucrados (Shirima et al., 2007). Las investigaciones sugieren que el uso de PDA tiene el potencial de reducir la carga logística, el costo y el índice de error asociados a la recopilación de información (Seebregts et al., 2009).

Los teléfonos móviles son usados cada vez más por los encuestadores para ingresar la información directamente. Tienen la ventaja de poder subir la información a un archivo de inmediato, evitando así el riesgo de perder la información si el teléfono se estropea o se pierde. También se puede hacer control de calidad y supervisar al encuestador en tiempo real, reduciendo con ello el riesgo de información fabricada (Tomlinson et al., 2009). Los teléfonos móviles ofrecen ventajas similares a los PDA en el sentido de reducir los errores que se producen durante el registro e ingreso de la información (Zhang et al., 2012). Se puede recurrir a diversos paquetes de *software* como SurveyToGo o Snap Surveys para configurar las encuestas para teléfonos móviles. Todos estos productos tienen modelos a diferentes precios y varían en cuanto al soporte a dispositivos específicos. Este campo está evolucionando vertiginosamente y por lo tanto es importante mantenerse al día con la oferta actual y los productos que resultan más convenientes.

Las encuestas suponen una serie de problemas logísticos para la EAH:

- **Falta de un marco muestral adecuado.** Un marco muestral es un listado más o menos acertado de la población de la cual ha de extraerse la muestra. La alternativa, cuando no existe otro marco muestral disponible, es utilizar muestras por conglomerados, pero esto aumenta el número de entrevistas necesarias por un factor de dos (nutrición) o cuatro (mortalidad). Ver la [Sección 12: Muestreo](#), para una discusión del efecto de diseño.
- **Tiempo.** El proceso de desarrollo y validación del cuestionario de la encuesta puede ser largo. Mandar hacer traducciones, capacitar a los encuestadores y llevar a cabo la encuesta son todas actividades que requieren tiempo, y las encuestas a menudo toman más tiempo que otros métodos en el sector de EAH.
- **Falta de una línea de base.** La utilidad de las encuestas es óptima cuando se cuenta con una línea de base con la cual comparar los resultados. Esto rara vez sucede en la EAH, pero las cosas están mejorando, a medida que se observa una mejora en las estadísticas nacionales de muchos países en desarrollo.
- **Falta de evaluadores con la capacidad necesaria.** La mayoría de los evaluadores en el sector de EAH tienen antecedentes académicos y profesionales cualitativos antes que cuantitativos y cuentan con experiencia limitada en la conducción de encuestas cuantitativas.

Diseñar un instrumento para una encuesta es una tarea compleja. Las preguntas de la encuesta, aunque más simples en ciertos sentidos que las preguntas de una entrevista semiestructurada, son más complejas en otros, especialmente porque el entrevistador no puede modificar la pregunta ni hacer preguntas de sondeo para que el entrevistado se explique en sus respuestas. Las encuestas estructuradas frente a frente se basan en el supuesto de que las preguntas se plantearán de la misma manera a cada entrevistado. Las preguntas de una encuesta tienen que ser cuidadosamente diseñadas y validadas antes de ponerse en práctica, ya que no pueden ser modificadas una vez que la encuesta entra a rodar.

La forma habitual de validar las preguntas de una encuesta es realizar una encuesta piloto. La forma en que se formulan las preguntas puede tener un efecto significativo en las respuestas, y una encuesta piloto puede sacar a la luz algunos de estos problemas. La monografía de Fowler (1995) sobre las preguntas de las encuestas es un documento especializado de utilidad.



#### Consejo útil

Si es preciso traducir la encuesta, es importante que una persona diferente la traduzca de vuelta a su idioma original.



### **Ejemplo de buena práctica: La experiencia de World Vision International (WVI) con herramientas de recolección de información**

Una vez que está lista la versión en inglés de las herramientas de recolección de información, se sigue el procedimiento siguiente:

1. Se contrata a un traductor de la región en la cual tendrá lugar la encuesta (con frecuencia alguien del personal local de WV, o en ocasiones también un consultor) para que las traduzca al idioma o dialecto local.
2. Las herramientas se traducen nuevamente a su idioma original (al inglés). Esto no siempre ha sido posible, pero es lo ideal.
3. Se lleva a cabo una sesión de capacitación con los encuestadores, en el transcurso de la cual estos verifican la exactitud de la traducción y se cercioran de que todos entiendan las preguntas de la misma manera. Se trata de un proceso muy iterativo.
4. Durante la capacitación de los encuestadores se lleva a cabo una prueba en el campo con las comunidades, en la que se utiliza el idioma de las herramientas de la encuesta (esto puede implicar cierta gimnasia logística si se requieren varios idiomas).
5. Luego de la validación de campo se revisan todas las respuestas a las encuestas y se hacen los cambios finales que resulten necesarios para las traducciones. Además, los encuestadores reciben retroalimentación sobre las respuestas que registraron y tienen una oportunidad final para revisar y aclarar cualesquier malentendidos.

Fuente: Kathy Duryee, WVI, comunicación personal, 2014

## **Encuestas en línea**

Las encuestas en línea son relativamente fáciles de armar y llevar a cabo. Han reemplazado en gran medida a las encuestas por correo y a algunas formas de encuestas telefónicas.

### **Características principales:**

- Ampliamente utilizadas en la EAH.
- Una herramienta sumamente útil para recabar la perspectiva de grupos extensos. Ideales para recoger la opinión del personal de la organización ejecutora o sus contrapartes, pero pueden servir también para recabar la opinión de la comunidad humanitaria más amplia.
- La encuesta puede ser contestada a distancia, independientemente de la zona horaria.
- Normalmente permanece abierta de dos a cuatro semanas.
- Puede recolectar información cualitativa y, de manera limitada, cuantitativa.
- Los encuestados pueden optar por participar o no, de modo tal que las tasas de respuesta pueden ser bastante bajas. Los encuestados normalmente se seleccionan en grupo, antes que aleatoriamente, lo cual

significa que los resultados no deben ser considerados representativos de la población más amplia.

- Son efectivas para identificar potenciales informantes clave – pregunten siempre a los encuestados si estarían dispuestos a ser contactados nuevamente para proporcionar información adicional.
- Los entrevistados pueden ser invitados pero básicamente se seleccionan a sí mismos – de modo tal que los hallazgos de las encuestas en línea no deben ser generalizados.
- Son económicas.
- Normalmente se reciben entre varias decenas y varios cientos de respuestas o más.
- Un problema ético es que en la práctica comercial regular se ofrece a los encuestados algún incentivo, por ejemplo la posibilidad de ganar un cupón o un dispositivo electrónico, pero esta práctica no se realiza en la EAH.



#### Consejo útil

- Limiten el número de preguntas para elevar la tasa de respuesta.
- Al inicio de la encuesta, especifiquen el número de preguntas o cuánto tiempo debería tomar responderlas.
- Envíen varios recordatorios con el fin de incrementar la tasa de respuesta. Si envían invitaciones individualizadas, la página de la encuesta puede rastrear quiénes ya han enviado sus respuestas y evitar que sigan recibiendo recordatorios.
- Realizar una encuesta en línea durante la fase inicial les permitirá identificar potenciales informantes clave, con opiniones fuertes o análisis bien desarrollados.
- Traducir la encuesta a los idiomas apropiados puede fortalecer la participación del personal nacional.
- Dillman et al. (2009) es la referencia habitual para encuestas en línea.

### Ejemplos del uso de encuestas en línea

- El estudio de OCAH, *Stay and Deliver*, sobre buenas prácticas en la entrega de la asistencia humanitaria en diferentes entornos de seguridad y riesgo aplicó una encuesta en línea en inglés, francés, español y árabe para recoger las opiniones de 1,148 empleados locales (Egeland et al., 2011).
- La evaluación de la respuesta de ACNUR al ingreso de refugiados a Líbano y Jordania utilizó una encuesta en línea para validar los hallazgos de la evaluación (Hidalgo et al., 2015).
- La evaluación del proyecto de desarrollo de capacidades ante emergencias (DCE) (Ky Luu et al., 2014).



### Consejo útil

Averigüen si una organización realiza encuestas de salida del personal – estas pueden ser útiles para identificar temas clave para una evaluación. Sin embargo, debe respetarse la confidencialidad.

La FICR ha empleado encuestas de salida del personal y las ha configurado para operaciones de emergencia a gran escala:

Hicimos eso para Haití, y también para Filipinas. Durante las primeras semanas después del desastre planteamos un número limitado de preguntas a nuestro personal saliente: sus percepciones, si cumplimos o no con nuestros principios fundamentales, qué tan bien estamos incorporando los pareceres y opiniones de la población afectada. Dejamos que la encuesta siguiera activa durante varios meses. La idea es que, para cuando hagamos una evaluación en tiempo real o una evaluación humanitaria, podamos transmitir a los evaluadores una buena cantidad de apreciaciones del personal que participó en la operación; muchas veces, esa gente también se ha ido. Es bueno que sea anónimo y definitivamente es una práctica que en mi opinión fue sumamente provechosa. (Josse Gillijns, comunicación personal, 2015)

## 13.5 Observación

La observación es una técnica valiosa en la evaluación y constituye una metodología medular de la EAH porque puede ayudar al evaluador a asimilar mejor el contexto. La observación puede ser estructurada o no estructurada.

### Observación estructurada

En la observación estructurada el observador utiliza un formulario para registrar la observación. Ha sido poco usada en la EAH pero existe espacio para incrementar su uso.

Por ejemplo, si una organización ha estado capacitando a las familias en la recolección de agua segura, podría ubicar a un observador al pie de la bomba de agua y este se encargaría de registrar las conductas para evaluar la eficacia de la capacitación. El observador dispondría de un formulario con las inscripciones que se aprecian a continuación en la primera columna, así como columnas subsiguientes para cada persona que acudiese al pozo a sacar agua.

Si el usuario observa la conducta en cuestión, se marca la columna correspondiente a dicho usuario en el formulario.

Usuario No.:	1	2	3	4	5	6	7	8	9	...
Elemento de buena práctica										
Trajo consigo un contenedor sellado.										
Abrió el contenedor sellado en el pozo.										
Mantuvo limpia la tapa.										
Enjuagó el contenedor.										
Echó el agua del enjuague en el desagüe.										
No provocó excesivas salpicaduras.										
Selló el contenedor antes de retirarse.										
Nota No. (para observaciones especiales):										
<b>Notas:</b>										

Algunos sistemas de observación estructurada pueden ser complejos,<sup>6</sup> pero otros pueden ser relativamente simples. Bentley et al. (1994) han elaborado una excelente guía sobre esta técnica, que incluye varios formularios de muestra así como orientaciones para preparar el formulario adecuado para cada situación. Aunque la observación estructurada se utiliza más ampliamente en saneamiento o educación, existen muchas opciones para usarla en la EAH, por ejemplo:

- Encuestas sobre el uso de servicios públicos tales como puntos de agua, letrinas comunitarias, etc.;
- Trato diferenciado dispensado a los hombres y las mujeres, los niños y las niñas, en puntos de servicio tales como centros de distribución o centros de salud;

- Atención provista por las oficinas de reclamaciones a diferentes clientes y tipos de reclamos;
- Nivel de atención e interacción durante las reuniones de coordinación.

No todos los ejercicios de observación estructurada requieren que la información sea codificada. En algunos casos el encuestador simplemente cuenta el número de personas que realizan una actividad en particular en un espacio de tiempo en particular. El encuestador no necesita tener ninguna habilidad específica, salvo perseverar en una tarea un poco tediosa. Se puede desarrollar un formulario de observación estructurada para cualquier fin – examinar las interacciones en las asambleas comunitarias, por ejemplo – dependiendo de las preguntas de evaluación.

Si los participantes son conscientes de la presencia del observador, pueden alterar su comportamiento. En Bangladesh, Ram et al. (2010) encontraron que la presencia de observadores aumentaba la frecuencia del lavado de manos. Se propusieron medir esto insertando sensores de movimiento en las barras de jabón. El jabón se movía significativamente en los días en que el lavado de manos era observado, y en el 22 por ciento de los hogares no se movió en absoluto los demás días.

## Observación no estructurada



### Consejo útil

De ser posible, observen a la comunidad antes de enfrascarse en entrevistas individuales o grupales con la población afectada. De esta manera podrán obtener información que les permita hacer las preguntas de sondeo adecuadas y promover un intercambio más abierto.

El equipo de evaluación de la respuesta al tsunami de Solidaridad Suiza realizó caminatas sistemáticas, recorriendo cada lugar de la evaluación al inicio de la investigación. El equipo tomó apuntes individualmente y luego los comparó, al final del día. Algunos miembros del equipo tenían experiencia previa en algunas de las ubicaciones, y esto configuró una importante base para hacer una comparación con las condiciones que encontraron (Ferf y Fabbri, 2014: 19).

En la observación no estructurada los evaluadores hacen observaciones mientras realizan caminatas transversales o simplemente recorren el lugar. Lee (2000) identifica cinco tipos de observación simple:

- **Señales físicas externas.** ¿Cómo está vestida la gente? ¿Hay signos exteriores de riqueza o pobreza? ¿O ambos? Un ejemplo de esto es una evaluación en Angola donde los líderes de la comunidad insistieron en que estaban pasando hambre debido a una reducción de las raciones pero se podían ver sacos de granos en muchos hogares.
- **Movimiento expresivo.** ¿Los niños están correteando y jugando? Un ejemplo de esto ocurrió en una evaluación en Malawi, donde una aldea en particular no recibió al equipo ejecutor de la organización con cantos y bailes, como era la costumbre. Era evidente que la aldea no estaba contenta con el equipo ejecutor.
- **La ubicación física se refiere a la organización del espacio por las personas.** ¿Hay evidencia de espacios compartidos o las personas se mantienen apartadas? ¿Se movilizan las personas por las carreteras? Desde luego algo de esto puede ser cultural pero también puede ser indicativo de temas subyacentes.
- **Conversaciones sobre el terreno.** Estas pueden consistir de intercambios breves que los evaluadores sostienen con los pobladores durante sus recorridos a pie por el lugar. ¿Se queja la gente? ¿Qué inquietudes manifiestan?
- **Conductas horarias.** Estas son más fáciles de capturar mediante el sistema de observación estructurada, pero igualmente es posible que el evaluador advierta cosas como personas que se desplazan desde un punto de distribución después del anochecer, lo cual sugeriría que la distribución no está bien organizada.



#### Consejo útil

Lleven una cámara fotográfica consigo siempre que sea apropiado. Las fotos son mucho más poderosas que las descripciones verbales de las observaciones. Fotografíen observaciones interesantes e incorporen las fotos en su informe.

El uso de material fílmico en las observaciones y para registrar las entrevistas también se está difundiendo. Ver la discusión sobre consentimiento informado y confidencialidad en la [Sección 14, Involucrando a la población afectada en su evaluación.](#)



#### **Ejemplo de buena práctica: Uso de fotos para reforzar mensajes clave**

En la respuesta a la crisis de Kosovo, una organización con considerable experiencia en el sector de agua potable y saneamiento ejecutó una serie de proyectos que no estuvieron a la altura de su desempeño habitual. Durante la reunión para informar a la organización sobre el trabajo, la sugerencia de que la organización no se había desempeñado con la eficiencia habitual fue rotundamente rechazada por el personal. Sin embargo, el ambiente cambió cuando el evaluador les mostró una serie de fotografías: por ejemplo una foto de una cisterna flexible de cinco toneladas en equilibrio precario en un entorno lleno de niños, para ilustrar la negligencia en la observación de las normas de seguridad y las políticas institucionales (Wiles et al., 2000).

## 13.6 Mediciones discretas

Las mediciones discretas tienen por objeto evitar interferir en la vida de la población afectada (Webb et al., 1966; Lee, 2000; Gray, 2014). Como señala Patton (2014), el objetivo de las mediciones discretas es evitar que la población afectada reaccione ante el intento de medir algo.

Internet ofrece una amplia gama de mediciones discretas, especialmente para el seguimiento de visitas a páginas web, búsquedas y medios sociales. Los temas pueden abarcar desde trivialidades como averiguar cómo un desliz de vestimenta afecta el interés en las celebridades (Pfister, 2011) hasta la frecuencia de búsquedas en Google del término “terremoto”. Desde luego que estos métodos son discretos, pero esta sección se refiere a mediciones observacionales no reactivas antes que a un análisis de contenidos, trátense de documentos convencionales, internet u otros recursos.



#### **Ejemplo de buena práctica: Uso de medios sociales por la FICR como medición discreta de la respuesta al tifón Haiyan en Filipinas**

En los tiempos que corren, cada desastre genera un *hashtag*, y en ocasiones más de uno, y al analizar el tráfico en Twitter o Instagram... podemos encontrar información que puede ser validada por el personal o por otras entrevistas. Se trata de una herramienta interesante para triangular algunos de los hallazgos que resultarían de una evaluación humanitaria.

(Josse Gillijns, comunicación personal, 2015)

Webb et al. (1966) sugirieron que se pueden hacer mediciones discretas de las señales de desgaste o acumulación.

El clásico ejemplo de mediciones basadas en el desgaste, citado por Webb et al., fue la observación del desgaste de las alfombras como un indicativo del interés de los visitantes en las diferentes exposiciones de un museo. Otro ejemplo de una medición basada en el desgaste fue el seguimiento realizado por Patton del consumo de café durante una capacitación de una semana de duración, para inferir qué sesiones eran las menos interesantes (Patton, 2014). En una situación de EAH, el desgaste de diferentes senderos puede sugerir qué tan intensamente se utilizan distintos puntos de agua y así sucesivamente.

Un ejemplo de una medición basada en la acumulación fue el estudio de Blake (1981) sobre el graffiti en Hawai como una medición discreta de las relaciones étnicas. En una situación de EAH, un ejemplo de una medición basada en la acumulación sería la tasa a la que se saturan las letrinas de un pozo, o el ritmo al que se genera la basura.

Las mediciones discretas son muy poderosas, ya que evitan el riesgo de que el acto de la medición afecte la medición en sí. Aunque existen algunos ejemplos de mediciones discretas en la gestión de la acción humanitaria, hay muy pocos ejemplos en la EAH propiamente dicha. Un ejemplo de una medición discreta a nivel de gestión fue la medición de la mortalidad con base en la entrega de mortajas (para envolver los cadáveres para enterrarlos) entre los refugiados de Tanzania Occidental. Estas se entregaban gratuitamente a pedido del usuario pero los refugiados eran tan supersticiosos que no pedían una mortaja para ningún otro fin que no fuera un funeral. Se trata de una medida erosiva, ya que agotó las existencias de las mortajas.

Un ejemplo del uso de una medición discreta de crecimiento en una evaluación fue el de la ETR interinstitucional de la respuesta a las inundaciones de Mozambique en 2007 (Cosgrave et al., 2007). Aquí, la medición del crecimiento se hizo del césped que creció en los albergues temporales que los aldeanos desplazados habían construido. Esto representó una fuerte evidencia de que los habitantes habían retornado a sus aldeas y volvían al albergue temporal únicamente para las distribuciones de la asistencia. Un ejemplo de una medición de disminución se dio en un campamento para refugiados donde el camino al inodoro del mercado mostraba un gran desgaste, mas no así el sendero a la estación de lavado de manos que quedaba al lado.

La evaluación de la respuesta rápida a los desplazamientos de población en la República Democrática del Congo estableció que la recolección de césped por los desplazados internos para la construcción de albergues temporales era una medida discreta de las demoras en la distribución de láminas de plástico (Baker et al., 2013: 22).

## 13.7 Metodologías orientadas al aprendizaje

Las metodologías señaladas aquí son especialmente apropiadas para evaluaciones orientadas al aprendizaje. Suelen ser participativas y deberían involucrar a todos los que necesitan o desean participar del aprendizaje. Las metodologías deben conducir a la reflexión y a la admisión de errores así como de éxitos. Este listado dista de ser exhaustivo.

En la cultura activista de las operaciones humanitarias, a menudo se afirma que no hay tiempo para ejercicios participativos y reflexivos. Sin embargo, solo una pequeña porción de la acción humanitaria es tan crítica en términos de tiempo, que dedicar un espacio a la evaluación la perjudicaría. Si se planean y facilitan debidamente, las actividades de aprendizaje pueden ser ejecutadas en un par de horas, o en un día, y el valor de hacer un alto en el camino para reflexionar y aprender colectivamente amerita la inversión de tiempo.

Después de identificar quién es el foco del aprendizaje – por ejemplo, miembros específicos del personal, contrapartes o miembros de la población afectada – es preciso diseñar procesos que les hagan reflexionar. Estos incluyen los siguientes:

- Crear un espacio seguro para la reflexión y el debate, en donde los participantes sientan que pueden hablar libremente y admitir errores;
- Hacer preguntas simples, abiertas, que suele ser la estrategia más efectiva para estimular la reflexión (ver “Questioning and Listening” [*Preguntar y escuchar*] en Chapman et al., 2005);
- Utilizar procesos participativos y creativos;
- Cerciorarse de que las enseñanzas aprendidas estén debidamente documentadas;
- Proveer una buena facilitación;
- Alentar a los participantes de los ejercicios de reflexión y aprendizaje a apropiarse de los hallazgos de la evaluación;
- Incorporar un proceso de seguimiento.

### Narración de historias

Contar historias es una forma natural de comunicación. Focalizar y facilitar este proceso puede ser una opción creativa para facilitar el aprendizaje, especialmente cuando va acompañado de un proceso de preguntas y reflexión encaminado a profundizar el análisis. Esta sección describe dos formas de aprovechar la narración de historias para fines de aprendizaje: el uso de metáforas y la técnica del cambio más significativo (CMS).

11 / Diseños de evaluación para responder las preguntas de evaluación 192 - 213

12 / Muestreo 214 - 230

13 / Metodologías de campo 231 - 267

14 / Involucrando a la población afectada en su evaluación 268 - 282

15 / Acceso restringido 283 - 292

16 / Análisis 293 - 321



### **Ejemplo de buena práctica: Contando historias con Oxfam GB en Aceh**

Justo después del tsunami en Asia, Oxfam realizó una revisión interna orientada al aprendizaje en Aceh. Para facilitar el autoaprendizaje del personal y hacerlo más participativo, se utilizó un enfoque de narración de historias. Este fue facilitado por la asesora de Global MEAL para Oxfam GB, quien explicó el ejercicio de la siguiente manera: “Yo no quería [usar] la introducción habitual, ‘Hola, soy de Oxfam y hemos venido a evaluar el programa’, porque ya con eso estás introduciendo un sesgo. Le pedimos al personal que no usara ninguna prenda de vestir con el distintivo de Oxfam y los dejamos en las afueras de la ciudad. Les pedimos que caminaran por la comunidad, se sentaran al costado de alguna persona y simplemente le dijeran: ‘Cuénteme su historia’. Debían acomodarse en el asiento y escuchar, y luego reportarnos lo que habían encontrado. Fue brillante. El personal realmente lo disfrutó; aprendieron cosas que no sabían del programa: errores que habían cometido, vacíos, gente que había sido excluida, errores de focalización, etc... Sintieron que la experiencia les había reportado un gran beneficio, y pudimos formarnos un mejor juicio de lo que estaba pasando en la comunidad que si hubiéramos hecho una evaluación más tradicional, donde le decimos a la gente de dónde venimos y por qué estamos ahí.”

Fuente: Vivien Walden, Oxfam GB, comunicación personal, marzo de 2015

### **Uso de metáforas**

La mejor forma de facilitar el uso de metáforas es en el contexto de un taller, por ejemplo siguiendo los siguientes pasos:

1. Los participantes, ya sea individualmente o en grupos, dibujan imágenes (por ejemplo, de un árbol o un río) para representar su experiencia del programa bajo evaluación. Las diferentes partes del árbol o río representan diferentes partes del programa.
2. Los participantes cuentan una historia sobre el dibujo, sin interrupciones, a sus colegas.
3. Los que están escuchando la historia hacen preguntas acerca de ella con la finalidad de profundizar el análisis y aprender, y ofrecer sus propias reflexiones.
4. Al final de este proceso, el grupo elabora un listado de los principales logros, desafíos y enseñanzas aprendidas de la narración de historias acerca de su experiencia común del programa.
5. Este ejercicio puede enfocarse en el futuro, realizando las siguientes preguntas al final:
  - a. ¿Cómo podemos reforzar lo que ha funcionado bien?
  - b. ¿Cómo podemos afrontar los desafíos?
  - c. ¿Cómo vamos a aplicar lo que hemos aprendido aquí?



### Ejemplo de buena práctica: Uso de metáforas y narración de historias en una revisión

En la revisión de Tearfund de la iniciativa de desarrollo de capacidades de gestión de desastres se utilizó la metáfora del árbol como base para un ejercicio de narración de historias. En un taller de seguimiento y aprendizaje, se le pidió a cada contraparte de Tearfund que dibujara un árbol.

- Las raíces representaban los valores y principios que sostenían el programa.
- El tronco representaba a las contrapartes (organizaciones y personas).
- Las ramas representaban las actividades.
- Las frutas representaban los logros del programa.
- Las hojas representaban las enseñanzas aprendidas.
- Las ramas rotas que yacían en el suelo representaban los desafíos internos.
- Las nubes en el cielo representaban los desafíos externos.
- Los brotes representaban las actividades planificadas pero aún no ejecutadas.
- La narración de historias fue seguida por una revisión del aprendizaje.

Pese a cierto escepticismo inicial, los participantes manifestaron que había sido un proceso participativo, visual y creativo que había invitado a la reflexión y al aprendizaje; llevado a la toma de conciencia colectiva de los éxitos, desafíos y enseñanzas aprendidas; y ofrecido una fotografía instantánea de todo el programa. Además, reconocieron el desafío de traducir lo aprendido en acciones concretas.

### La técnica del cambio más significativo (CMS)

La técnica CMS es un enfoque participativo que puede usarse con el personal o con miembros de la comunidad local. Cuando se emplea con la población afectada, puede enfocarse en los efectos o el impacto. Va evolucionando en forma ascendente desde el nivel de campo, utilizando historias que capturan lo que ocurrió, cuándo y cómo, y por qué era importante. Esta metodología fue utilizada en la evaluación de la FICR de la respuesta al terremoto de 2007 en el Perú: se pidió a los participantes que identificaran los cambios más significativos que hubieran ocurrido como resultado de la intervención de la Cruz Roja Peruana (Martínez, 2009).

Los pasos son los siguientes:

1. Se recolectan historias a nivel de campo en respuesta a dos preguntas. En su opinión, durante el mes pasado [u otro período de tiempo], ¿cuál ha sido el cambio más significativo que ha tenido lugar en las vidas de las personas que participan en el proyecto? ¿Por qué fue el cambio más significativo?

2. Las más significativas de estas historias son seleccionadas por un panel de partes interesadas o miembros del personal.
3. Una vez que los cambios han sido registrados, se leen las historias en voz alta, a menudo en el contexto de un taller, y los participantes debaten y reflexionan sobre el valor de los cambios reportados.

Ver consejos sobre el uso de esta técnica en Davies y Dart (2005).

## Talleres

Los talleres con el personal de la organización son una característica común de la EAH. Los talleres a menudo se llevan a cabo:

- A inicio del trabajo de campo. Esto permite que el equipo de evaluación se presente, presente el [informe inicial](#) y reciba retroalimentación del personal de campo. Es también una manera eficiente de concertar entrevistas posteriores con informantes clave.
- Al final del trabajo de campo. Esto permite que el equipo de campo informe sobre lo que han descubierto y valide sus hallazgos y conclusiones con los participantes del taller.

El equipo de evaluación puede usar ambos talleres para recolectar información, identificar aspectos clave y afinar sus hallazgos, ya sea de la fase inicial o del trabajo de campo. Como se indicó en la [Sección 5, Enmarcando su evaluación](#), los talleres pueden usarse para desarrollar o validar la TdC y para establecer qué califica como adecuado, bueno y excelente para elaborar una rúbrica evaluativa (ver la [Sección 10, Métodos documentales](#)).

El informe de la ETR de Oxfam de su respuesta a la sequía en Kenia reflejaba “la retroalimentación recibida en los talleres realizados con las contrapartes y el personal para informar sobre el trabajo de campo” (Murphy et al., 2011: 2). Se trata de un enfoque bastante común. Los talleres de puesta en marcha son menos comunes que los talleres para informar sobre el trabajo de campo, pero tienen el potencial de incrementar la apropiación de la evaluación entre las partes interesadas. Además, son útiles para organizar las entrevistas y la logística de las visitas de campo.



### Consejo útil

En el taller de puesta en marcha, indaguen qué preguntas les gustaría a los participantes que respondiera la evaluación. No será posible incidir en todas, pero incluso esta discusión puede aclarar lo que la evaluación realmente va a hacer y reforzar su aceptación.

Los talleres también pueden usarse con fines más amplios. En la evaluación de impacto de las transferencias de dinero en efectivo en Sumatra Occidental se organizó un taller de puesta en marcha así como un taller final de aprendizaje para presentar los hallazgos de la evaluación. Dicho taller de aprendizaje representó además una oportunidad para que las partes interesadas compartieran prácticas idóneas y desarrollaran algunas medidas prácticas para incidir frente a las entidades donantes con el objetivo de lograr su apoyo a los enfoques de transferencias de dinero (Aspin, 2010). Durante la ETR de tres meses de duración de la respuesta al terremoto de Haití, las ONG locales organizaron un taller con los evaluadores para empezar una discusión sobre cómo mejorar la colaboración con los actores internacionales (Grünewald et al., 2010: 18).

### Características principales:

- Ampliamente utilizados en la EAH, normalmente para informar sobre los hallazgos al término del trabajo de campo.
- Suelen ser celebrados con un grupo de personas versadas en el tema que se está evaluando.
- A menudo utilizados para desarrollar o validar la TdC o para validar el plan de evaluación o las conclusiones de la evaluación.
- Generan información sobre la validez de la propuesta.
- Pueden usarse para formular las recomendaciones con base en los hallazgos y conclusiones de la evaluación.
- Se necesita un local para el taller, sillas, papelógrafos, proyector, etc.
- Tradicionalmente se ofrece una comida.
- Las organizaciones participantes generalmente se seleccionan intencionalmente, pero la organización decide quiénes asistirán. El principal problema con los talleres es que a menudo la mayoría de los potenciales participantes más útiles están muy ocupados para asistir.
- Se trata de una metodología de costo medio.
- Normalmente se celebran uno o dos talleres, pero en algunos casos puntuales puede celebrarse un número mayor para una evaluación pequeña de dos semanas de duración.

Los talleres pueden ser menos útiles si el mismo evento congrega a grupos con intereses muy diversos o con diferencias importantes a nivel socioeconómico o de autoridad. Esto dificulta la construcción de consensos en torno a temas específicos.



### Consejo útil

Consideren organizar talleres en cascada a distintos niveles. Una serie de talleres en cascada convierte las conclusiones de un taller en el tema del siguiente. Por ejemplo, un taller inicial de un día de duración para el personal de campo o las contrapartes podría identificar los temas que requieren solución a ese nivel. En un taller para gerentes de nivel medio celebrado al día siguiente se podrían considerar los temas identificados e identificar algunos de los problemas subyacentes responsables de generarlos. Luego, se podría sugerir un número de opciones para abordar dichos problemas. Finalmente, un taller de medio día de duración para la alta dirección podría revisar los resultados de los dos primeros talleres y decidir qué medidas tomar para solucionar los problemas.

Ver la Sección 13: Metodologías de campo, para un ejemplo del uso de talleres para elaborar las recomendaciones.

### Herramientas de aprendizaje en los talleres

Los talleres pueden servir para capturar el conocimiento de los participantes y también para ayudar a sacar a la superficie el conocimiento de las partes interesadas existentes y darles una oportunidad para verbalizarlo y analizarlo. Un formato establecido para talleres de aprendizaje es el examen retrospectivo, que se analiza más adelante, pero se pueden usar muchas otras herramientas en los talleres para promover el aprendizaje, entre ellas las siguientes:

- **Dotmocracy o puntocracia (formato 'qué opinas')**: en la que los participantes indican que están de acuerdo con diferentes afirmaciones colocando puntos sobre ellas ([www.dotmocracy.org](http://www.dotmocracy.org)).
- **Hojas de calificación de ideas**: en la que cualquier participante puede generar una idea y luego los demás participantes pueden estar de acuerdo con ella o no ([www.ideaatingsheets.org](http://www.ideaatingsheets.org)). Las hojas de calificación de ideas resuelven algunos de los problemas de la puntocracia.
- **Minirrúbricas**: en la que los participantes utilizan puntos de diferentes colores para revelar su opinión acerca del valor u otros aspectos o partes de una intervención. Pueden utilizarse también para presentar rápidamente los resultados de una evaluación (Davison, 2014).
- **Categorización jerárquica de cartillas**: una opción participativa que consiste en categorizar diferentes cartillas y ha sido diseñada para revelar la forma en que las personas categorizan y jerarquizan diferentes fenómenos. Similar a la jerarquización que tiene lugar como parte de un diagnóstico rápido participativo (DRP) pero debe usarse con grupos que sepan leer y escribir.<sup>7</sup>
- **Rich Pictures**: herramienta que se utiliza para explorar, reconocer y definir una situación mediante el uso de diagramas con el fin de crear un modelo mental preliminar.<sup>8</sup>
- **Diagramas**: en los que se pide a las partes interesadas que representen las relaciones en forma de diagrama o que elaboren árboles o mapas de problemas.

Existe un número mucho mayor de herramientas adicionales que han sido desarrolladas para el DRP o para fines de capacitación y pueden ser utilizadas para recolectar información con diferentes grupos de participantes.

### Talleres de examen retrospectivo

Los exámenes retrospectivos pueden usarse como un enfoque de aprendizaje alternativo a la evaluación y también como una metodología para responder a preguntas de evaluación.



#### **Definición: Examen retrospectivo**

Un examen retrospectivo es una discusión estructurada de una intervención humanitaria, que le permite al equipo considerar y reflexionar acerca de lo que sucedió, por qué sucedió y cómo sostener las fortalezas y mejorar las debilidades.

El examen retrospectivo suele ser un proceso facilitado que involucra a todos los que han sido identificados como el foco de aprendizaje. Se lleva a cabo en el contexto de un taller. Una atmósfera franca y abierta, que incentive la confianza entre los participantes, es fundamental para el éxito del examen retrospectivo. El principio general es que no hay atribución ni castigo. Un facilitador neutral y objetivo es fundamental para asegurar que la discusión permanezca enfocada en los temas, mantenga una orientación positiva y no se deteriore, cayendo en la autojustificación o la culpabilización de los demás.

El examen debe tratar de responder las siguientes preguntas:

- ¿Qué se esperaba que sucediera?
- ¿Qué sucedió en realidad?
- ¿Qué funcionó bien y por qué?
- ¿Qué podría haber funcionado mejor y por qué?

El examen retrospectivo incluye los pasos que se establecen a continuación. Para cada paso, los participantes individuales podrían escribir sus respuestas en notas adhesivas, que el facilitador reuniría luego en conglomerados para reflejar temas particulares, o bien el grupo podría responder cada pregunta colectivamente y el facilitador podría escribir las respuestas en un papelógrafo.

- Los participantes, ya sea individualmente o en parejas, escriben su percepción del objetivo o intención de la intervención (dado que la acción humanitaria a menudo se lleva a cabo en entornos caóticos y los planes pueden resultar obsoletos rápidamente, se trata de un buen sistema para averiguar si el objetivo está claro y es compartido por todo el mundo, y redefinirlo de ser necesario).

- Acto seguido, los participantes escriben lo que realmente sucedió – posiblemente trabajando en grupo en la elaboración de una línea de tiempo de los acontecimientos más importantes y los cambios en la situación o en el programa a través del tiempo. Colocar cartillas en una pizarra o una pared para mostrar la línea de tiempo puede ser útil para este ejercicio.
- A continuación, el grupo intenta responder a dos preguntas: ¿Qué funcionó bien? ¿Qué podría haber funcionado mejor?
- A esto le sigue una discusión de lo que podría hacerse de manera diferente la próxima vez, al final de la cual el facilitador hace un resumen de todas las enseñanzas que han emergido e invita a los participantes a votar las tres enseñanzas que consideran más importantes.

Las enseñanzas centrales aprendidas, y cualesquier recomendaciones plausibles de ser puestas en práctica, son documentadas y circuladas entre todos los participantes. Se puede acordar un plazo, por ejemplo de seis meses a un año, para evaluar los avances en la implementación de las recomendaciones.

#### **Características principales:**

- En la EAH, los exámenes retrospectivos típicamente se utilizan cuando una organización desea extraer lecciones de sus operaciones.
- No suelen durar más de medio día o un día.
- Se requiere un facilitador y una persona adicional para tomar apuntes.
- Es preciso respetar los principios de no atribución y no sanción – la organización debe tener una cultura institucional que permita identificar los problemas sin buscar chivos expiatorios después.
- Son efectivos para ayudar al equipo de evaluación a identificar las enseñanzas clave de la operación.
- Necesitan un local apropiado, y el personal clave debe tener tiempo para participar.
- Los participantes son seleccionados intencionalmente.
- Son relativamente económicos.
- Normalmente solo se lleva a cabo un examen retrospectivo por ubicación, de modo tal que para una evaluación de varias ubicaciones posiblemente se requiera más de un examen retrospectivo.

La guía de USAID (2006) ofrece detalles del proceso.



**Ejemplo de buena práctica: Examen retrospectivo interinstitucional de la respuesta al terremoto y los tsunamis de 2004 en el Océano Índico**

Este examen, realizado por el personal (Baker, 2005) de cinco ONG de cuatro países afectados por el tsunami, duró dos días y se identificaron tres enseñanzas principales:

- La necesidad de un análisis socioeconómico temprano para asistir con la programación y el seguimiento de los programas, en el caso de evaluaciones conjuntas rápidas;
- El rol central que cumplen la consulta con la comunidad y su participación;
- La importancia de planificar medidas de preparación, en especial la necesidad de desarrollar capacidades locales de respuesta a situaciones de emergencia.

# 14 / Involucrando a la población afectada en su evaluación

Las organizaciones humanitarias están prestando renovada atención al fortalecimiento de sus acciones de rendición de cuentas frente a las poblaciones afectadas. Se trata de un compromiso central del Programa de Transformación, que tiene por objeto mejorar la eficacia de la acción humanitaria. Existe un creciente interés en robustecer el enfoque de “comunicación con las comunidades” (CC).<sup>9</sup>

El informe “Estado del sistema humanitario” de 2015 reportó:

...el aumento de la sensibilización sobre el terreno acerca de la importancia de involucrar a las personas afectadas... de modo tal que las poblaciones afectadas por los conflictos y desastres no se consideren simplemente “destinatarios” y que las intervenciones se diseñen con un mayor enfoque en sus necesidades y preferencias... Sin embargo, los avances en la rendición de cuentas frente a los beneficiarios de la ayuda han sido más retóricos que reales. (ALNAP, 2015: 72; ver también Ashdown, 2011; y Brown y Donini, 2014)

¿Qué significa esto para la EAH? ¿Cómo involucrar más efectivamente a las poblaciones afectadas en las evaluaciones, e igualmente en las fases de ejecución previas?

Involucrar a las poblaciones afectadas ha sido un punto débil de la EAH, a pesar de varios intentos de fortalecer la rendición de cuentas frente a ese grupo crucial – después de todo, se trata de las personas en cuyo nombre se emprende la acción humanitaria.<sup>10</sup> Si nos tomamos en serio lo de escuchar a la población afectada, es esencial involucrarla en el transcurso de una evaluación.

## 14.1 Involucrando a la población afectada

Solo involucrando a la población afectada y escuchando sus perspectivas y puntos de vista podemos saber si los programas y proyectos humanitarios han sido en efecto apropiados para satisfacer sus necesidades, y por lo tanto podemos cotejar las acciones con los indispensables criterios del CAD de la OCDE en materia de pertinencia. Las personas afectadas por una crisis son los jueces más idóneos de la eficacia del trabajo humanitario. Si no escuchamos sus opiniones, no podremos mejorar la calidad de la acción humanitaria. Esto significa incorporar la consulta sobre la calidad de las intervenciones en las evaluaciones.

La Asociación Internacional para la Rendición de Cuentas en el Ámbito Humanitario (HAP), actualmente Alianza CHS (Normas Humanitarias Esenciales), lleva muchos años trabajando en el tema de involucrar a las poblaciones afectadas, desarrollando una normatividad que subraya que escuchar lo que las personas afectadas por una crisis tienen que decir es algo que debe hacerse en todas las etapas del ciclo del proyecto, desde las fases de planificación y ejecución hasta el seguimiento y evaluación de los programas humanitarios (HAP, 2013). Gran parte de lo aprendido por la Alianza CHS ha sido incorporado en las Normas Humanitarias Esenciales (2015).

En la práctica, las organizaciones tienen diferentes razones para involucrar a la población afectada en sus programas humanitarios, aunque esto no siempre sea explícito. Brown y Donini (2014: 20-21) identifican tres tipos de racionalidad para que las organizaciones involucren a las poblaciones afectadas:

1. **Racionalidad normativa o basada en valores:** las organizaciones consideran que es lo que se debe hacer, por ejemplo para cumplir con un deber moral o respetar los derechos fundamentales y la dignidad de las personas afectadas;
2. **Racionalidad instrumental:** porque incrementa la eficacia de los programas humanitarios, por ejemplo al recoger información para enriquecer las decisiones programáticas y satisfacer mejor las necesidades de los afectados por la crisis;
3. **Racionalidad emancipadora:** porque aborda desigualdades estructurales y las causas subyacentes de una crisis, por ejemplo al dar una voz e iniciativa a las poblaciones marginadas o, más ambiciosamente, transformar las estructuras y las dinámicas del poder.

La evaluación desempeña un papel vital en evaluar la eficacia de las organizaciones en lograr dicho involucramiento. Cuando una racionalidad se encuentra implícita, es posible que la evaluación deba explicitarla, por ejemplo a partir de las entrevistas durante la fase inicial. A medida que las organizaciones se esfuerzan por asumir una mayor rendición de cuentas frente a las poblaciones afectadas, por ejemplo como parte del Programa de Transformación, la evaluación se convierte en un instrumento vital para evaluar si lo han logrado o no desde el punto de vista de las poblaciones afectadas. El ejemplo de buena práctica que aparece en la página 270 sintetiza los hallazgos de una investigación efectuada en Filipinas para recabar las opiniones de las comunidades afectadas después del tifón Haiyan, a raíz de que muchas organizaciones humanitarias invirtieran tiempo y recursos en aumentar su rendición de cuentas frente a la población afectada.



**Ejemplo de buena práctica: Escuchando a la población afectada acerca de los esfuerzos de las organizaciones para rendirles cuentas**

Por medio de una investigación efectuada entre noviembre de 2014 y febrero de 2015 como parte del proyecto de Servicios Comunes *Pamati Kita* (“Escuchemos Juntos”), las comunidades afectadas por el tifón Haiyan fueron consultadas acerca de sus perspectivas y experiencias con relación a los esfuerzos de las organizaciones humanitarias para rendirles cuentas. La investigación adoptó un enfoque cualitativo, en consonancia con un proyecto centrado en la escucha activa. Las consultas revelaron que la población afectada tenía una marcada preferencia por la comunicación cara a cara, frente a medios de comunicación más tecnológicos como las líneas de mensajería electrónica directa, que habían sido usadas por muchas organizaciones. En términos generales, la población afectada describió su relación con las organizaciones humanitarias internacionales como bastante distante, en contraste con las perspectivas de las organizaciones, que creían haber sido mucho más accesibles.

Fuentes: Ong et al. (2015); Buchanan-Smith et al. (2015)

Las evaluaciones que han involucrado activamente a la población afectada suelen ser mucho más ricas que aquellas que no lo han hecho. Ver, por ejemplo, la evaluación conjunta de CARE y Save the Children tras el terremoto de Haití en 2010 como [ejemplo de buena práctica](#) a continuación. Solo en ciertos tipos de evaluaciones podría haber razones de peso para no involucrar a la población afectada, por ejemplo en una evaluación exclusivamente enfocada en los procesos institucionales antes que en la eficacia de un programa para aliviar el sufrimiento.



**Ejemplo de buena práctica: Método de impacto “poner a la gente primero”**

Nueve meses después del terremoto de Haití, CARE Internacional y Save the Children (SCF) encargaron una evaluación independiente conjunta de su respuesta humanitaria. Si bien la evaluación se basó en los criterios de evaluación del CAD de la OCDE y examinó temas transversales para evaluar los esfuerzos de ayuda hasta la fecha, también sirvió como una fotografía instantánea de la forma en que diferentes grupos representativos de la sociedad haitiana percibían las áreas de la respuesta humanitaria global en las que ambas organizaciones habían participado.





Esto se hizo utilizando la metodología de impacto conocida como “poner a las personas primero”, que efectivamente colocaba a los habitantes de Haití en el centro, y se trabajó a partir de esa experiencia para determinar la eficacia de esfuerzos institucionales más amplios, incluidos los de CARE y SCF. El personal nacional de ambas organizaciones recibió una capacitación de dos días para perfeccionar sus habilidades de comunicación, escucha y facilitación, y luego para organizar discusiones de grupos focales utilizando preguntas temáticas abiertas. Esto quería decir que los temas de la discusión no podían ser impuestos, extendiéndose más allá del trabajo y los proyectos de cualquiera de las dos organizaciones, con la finalidad de obtener información sobre las experiencias reales de la gente y por lo tanto ayudar a responder las preguntas de ambas organizaciones: “¿Estamos haciendo las cosas correctas?” y “¿Estamos haciendo las cosas correctamente?”.

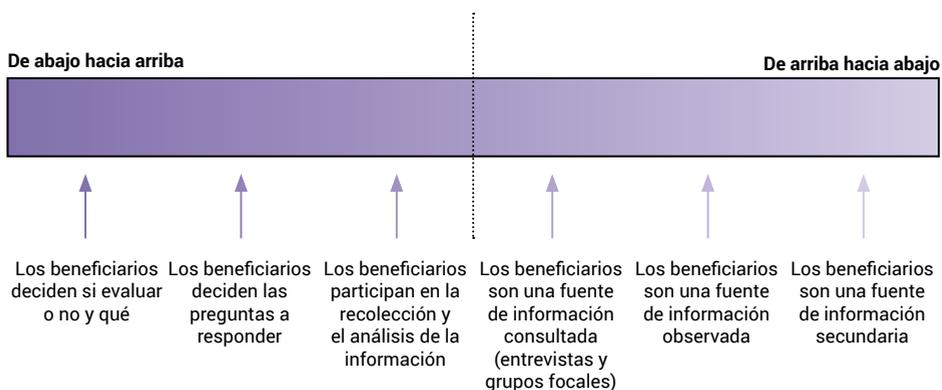
Fuente: O'Hagan et al. (2010)

Las evaluaciones de intervenciones humanitarias verdaderamente “de abajo hacia arriba” y participativas (ver la [Sección 4, Tipos de evaluación](#)) – en el extremo izquierdo del espectro en la [Figura 14.1](#) – son sumamente raras en el sector humanitario. Hacerlo implicaría involucrar a la población afectada en todos los aspectos de la evaluación, desde las fases de planificación y diseño, pasando por la recolección y el análisis de la información, la identificación de los hallazgos y la elaboración de las recomendaciones, hasta la [difusión](#).<sup>11</sup> Las secciones a continuación ofrecen orientación sobre las opciones más comunes para involucrar a la población afectada en la EAH.

## 14.2 Planificando involucrar a la población afectada

Rara vez se involucra a la población afectada desde el inicio de una evaluación, para determinar la necesidad de hacerlo así como sus potenciales alcances, aunque existen precedentes. Un ejemplo es el estudio de factibilidad realizado por OCHA para una evaluación de impacto conjunta (Beck, 2011). Para ello se requiere asignar tiempo para consultas *antes* de que inicie la evaluación.

Figura 14.1: Continuo participativo en la evaluación – ejemplo de la FICR



Fuente: FICR, Departamento de Planificación y Evaluación (2014)

Más a menudo, involucrar a la población afectada significa planificar consultas con ella durante el proceso de evaluación. Tal como se demuestra en el continuo participativo de la evaluación de la FICR, esto tiende a suceder en la mitad “de arriba hacia abajo” del espectro participativo, pero es importante pese a ello por las razones establecidas arriba. La entidad que encarga la evaluación debe planificar este tipo de consulta desde el principio, cerciorándose de que tanto los recursos disponibles como los plazos son adecuados para períodos de trabajo de campo sustanciales.

La consulta con la población afectada en la EAH ha sido débil, en general debido al poco tiempo y los limitados recursos que se destinan al trabajo de campo; esta suele ser la parte que se recorta cuando los plazos tienen que ser acortados y los presupuestos son ajustados. Y sin embargo, tal como lo subrayó un documento inicial de ALNAP y Groupe URD sobre la participación de las personas afectadas por una crisis en la acción humanitaria (ALNAP, 2003), se trata no solo de un conjunto de herramientas sino también de una actitud, en este caso una actitud que coloque el involucramiento y la consulta con la población afectada en el centro del proceso de evaluación y les asigne los recursos adecuados.

Los coordinadores de una evaluación deben cerciorarse asimismo de contratar evaluadores con el conjunto de capacidades y los conocimientos adecuados. Deben tener experiencia en la realización de consultas con las poblaciones afectadas, y algunos miembros del equipo deben tener conocimiento previo del contexto y entender las estructuras comunitarias y la dinámica del poder. Esto puede ser especialmente importante en una crisis ocasionada por un conflicto. Si hay temas de protección sensibles a tener en cuenta, es importante que los miembros del equipo de evaluación posean las capacidades y la experiencia necesarias para saber cómo afrontarlos.



### Consejo útil

Cuando contraten a los evaluadores para llevar a cabo una consulta con las poblaciones afectadas, además de entrevistarles y revisar sus hojas de vida, averigüen si han realizado trabajos similares anteriormente y, si procede, pídanles referencias.



### Consejo útil

Al planificar un proceso de consulta con la población afectada en contextos difíciles, por ejemplo debido a la falta de acceso y/o la situación de inseguridad, incorporen plazos de contingencia al cronograma de trabajo de la evaluación, en caso de que se produzcan retrasos en los viajes o deban hacerse arreglos alternativos.

## 14.3

### Diseñando la evaluación de modo tal que la población afectada se vea involucrada

El primer paso en el diseño de una evaluación consiste en identificar qué preguntas de evaluación de aquellas que figuran en los términos de referencia son pertinentes para consultarlas con la población afectada. Las preguntas acerca del impacto más amplio de la intervención humanitaria, y sobre su adecuación y pertinencia respecto a las necesidades, sin duda deben dirigirse a la población afectada. Pero las preguntas referidas a la eficiencia de los costos de la intervención, o a la gestión interna de la organización, probablemente sea irrelevante someterlas a consulta con la población afectada.

Segundo, es importante revisar lo que ya existe en términos de retroalimentación de la población afectada, especialmente en los casos en que haya habido un alto nivel de involucramiento a lo largo de la ejecución de la evaluación, por ejemplo a través de mecanismos de retroalimentación tales como líneas directas, seguimiento post distribución de la ayuda y consultas regulares con la comunidad, y en que la retroalimentación haya sido debidamente documentada. Por ejemplo, tras el paso del tifón Haiyan por Filipinas, algunas organizaciones como World Vision International (WVI) contaban con grandes bases de datos computarizadas, que registraban y analizaban la retroalimentación recibida, habiendo invertido en una serie de mecanismos diversos de rendición de cuentas.<sup>12</sup>

Tercero, es importante preguntarse qué sectores de la población afectada deben ser consultados, y por lo tanto cómo se debe hacer el desglose de la población. Asegúrense de consultar la guía de UNEG sobre la incorporación de los derechos humanos y la igualdad de género en las evaluaciones (2014). Un buen ejemplo es la evaluación de la política de transversalización del enfoque de edad, género y diversidad de ACNUR en Colombia (Mendoza y Thomas, 2009). El desglose de la información normalmente se hace para sacar a la luz diferentes experiencias con relación a la crisis

y la respuesta. Así, por ejemplo, hombres y mujeres, adultos mayores y jóvenes, grupos más acomodados y grupos pobres pueden haber experimentado la crisis y la respuesta de maneras diferentes. Tengan en cuenta asimismo las diferencias irreconciliables, especialmente en una crisis vinculada a un conflicto: por ejemplo, el origen étnico puede ser una importante barrera infranqueable en un conflicto, en el que diferentes grupos étnicos experimentan la crisis de maneras muy diferentes.

Cuarto, es preciso seleccionar las metodologías y el enfoque de muestreo adecuados para involucrar y consultar a la población. En la práctica suelen utilizarse metodologías cualitativas para involucrar a la población afectada. En ocasiones se ha consultado a la población afectada mediante “metodologías mixtas”, es decir, una combinación de encuestas cuantitativas formales y técnicas cualitativas, lo cual puede ser particularmente revelador si se hace bien, pero para lo cual se necesitan mayores recursos (Gayfer et al., 2014). Esto ha sido más común en el caso de evaluaciones conjuntas que han contado con abundante financiamiento, y de manera creciente también en evaluaciones de impacto. Tal como se describió en la [Sección 12, Muestreo](#), a menudo se recurre al muestreo intencional debido a limitaciones en términos de acceso restringido y dificultades para llegar a determinadas áreas geográficas o a ciertos sectores, lo cual dificulta enormemente la toma de una muestra aleatoria. En ese caso, el conocimiento local es clave para ponerse de acuerdo en torno al muestreo intencional.

## 14.4 Métodos para involucrar a la población afectada

El [Cuadro 14.1](#) contiene un resumen de cinco de las metodologías más comunes para consultar con la población afectada y las ventajas y desventajas de cada una. El diagnóstico rápido participativo (DRP) es también una metodología muy común y se aborda en la subsección siguiente.

Los grupos focales y las entrevistas grupales son particularmente eficaces para escuchar a la población afectada en el marco de discusiones más abiertas. Normalmente se guían por una lista de verificación de los temas a cubrir, al mismo tiempo que asignan el espacio necesario para ahondar en informaciones nuevas o inesperadas que salgan a la luz durante la discusión. Los grupos focales representan asimismo una oportunidad para consultar con diferentes grupos poblacionales por separado, por ejemplo mujeres y hombres, diferentes grupos de edad. Un grupo focal debe constar de seis a ocho personas pertenecientes a un grupo homogéneo de la población. En la práctica, rara vez es posible controlar el tamaño y la composición del grupo, que se puede convertir en una “entrevista grupal” de mayores dimensiones.

Por razones protocolares y de cortesía, puede ser apropiado empezar la consulta con una asamblea comunitaria, para explicar la evaluación. Esta es también una oportunidad para preguntar por la perspectiva general de la comunidad sobre el programa o proyecto, pero es probable que no todas las opiniones sean expresadas. Por ejemplo, las mujeres, los jóvenes y los grupos marginados pueden ser reacios a hablar en público.

**Cuadro 14.1:** Metodologías de consulta comunes con la población afectada

Metodología	Ventajas	Desventajas
Grupos focales y entrevistas grupales	Eficaces para consultar con un grupo en particular; Eficaces para discusiones abiertas.	Dificultad para controlar el tamaño y la composición del grupo; Requieren facilitación experimentada.
Asambleas comunales	Aprovechan estructuras ya existentes; Pueden involucrar a una gran cantidad de gente.	Pueden ser dominadas por ciertas personas o grupos; Dificultad para desagregar las respuestas.
Entrevistas personales y de hogares	Mayor facilidad para hacer preguntas sensibles; Los temas pueden ser abordados en mayor profundidad; Pueden ser ilustrativas.	Requieren más tiempo; Una muestra pequeña no permite hacer generalizaciones.
Entrevistas semiestructuradas con informantes clave	Sirven para obtener una buena apreciación general y puntos de vista a los que se puede hacer seguimiento en entrevistas grupales o personales.	Necesidad de estar atentos a posibles "guardianes" y a los potenciales sesgos de informantes clave.
Encuestas formales	Arrojan conjuntos de datos comparables y pueden ser más fáciles de generalizar; Información cuantitativa convincente para las instancias de decisión.	Requieren tiempo y son relativamente costosas; Pueden no ser viables en entornos de conflicto.

Las discusiones de grupos focales pueden ser convenientemente profundizadas mediante una serie de entrevistas personales o de hogares, con el fin de indagar sobre temas puntuales y/o hacer seguimiento de información sensible no apta para ser discutida en grupo. Pueden ser especialmente útiles como estudios de caso ilustrativos, demostrando el efecto a nivel de hogares de un aspecto en particular de una crisis y/o respuesta que hubiera emergido como un patrón común, por ejemplo en el marco de discusiones de grupos focales.

Las entrevistas con informantes clave son ampliamente utilizadas en la EAH como un medio para consultar con la población afectada. Los informantes clave suelen ser personas de la comunidad que están en posición de comentar la respuesta humanitaria en general y sus consecuencias, por ejemplo, quién recibió asistencia humanitaria y quién no y por qué. También es posible que puedan hablar acerca de algunos de los efectos no intencionales de la respuesta, sean positivos o negativos. Entre los informantes clave se encuentran los trabajadores de salud y los maestros, que tienen un contacto fluido con los miembros de la comunidad y están bien posicionados para observar los cambios y las tendencias así como qué grupos son más vulnerables y se encuentran más marginados, por ejemplo. Tras el tsunami en Tailandia, los evaluadores descubrieron que los monjes y policías eran valiosos informantes clave, a pesar no de haber recibido ayuda humanitaria.

Las encuestas formales también pueden usarse como un medio para consultar con la población afectada, aunque, como se describe en la Sección 13, Metodologías de campo, se rigen por preguntas predeterminadas y no tienen cabida para las discusiones abiertas.

11 / Diseños de evaluación para responder las preguntas de evaluación 192 - 213

12 / Muestreo 214- 230

13 / Metodologías de campo 231 - 267

14 / Involucrando a la población afectada en su evaluación 268- 282

15 / Acceso restringido 283 - 292

16 / Análisis 293 - 321



## Diagnóstico rápido participativo

### Definición: Diagnóstico rápido participativo

Las técnicas de diagnóstico rápido participativo o diagnóstico rural participativo (DRP) son un conjunto de metodologías que permiten que la población local enriquezca, analice y comparta su conocimiento y aprendizaje de formas que sean fácilmente entendibles para los evaluadores externos.

Si bien las técnicas de DRP fueron originalmente diseñadas para contextos de desarrollo, son muy apropiadas para involucrar a la población afectada.

### Características principales:

- Las metodologías de DRP generan información cualitativa principalmente.
- Se utiliza en la EAH para consultar con la población afectada.
- Los materiales requeridos dependen del método específico.
- Las ubicaciones se seleccionan intencionalmente.
- Los participantes pueden ser autoseleccionados o pertenecer a un grupo en particular.
- Puede realizarse como parte de una entrevista grupal.
- El número de sesiones de DRP puede oscilar entre unas pocas y varias decenas. Si constituye la principal técnica utilizada en una evaluación participativa, se requerirán más sesiones.



### Tener presente

El DRP es de bajo costo para los evaluadores pero exige una gran inversión de tiempo de los participantes. ¿Es esto apropiado o razonable en esta fase de la crisis y/o en el contexto actual?

Un punto fuerte de la mayoría de las técnicas de DRP es que son de base comunitaria y por lo tanto permiten a los evaluadores capturar el saber comunal, y no solo de uno o dos de sus miembros. La mayoría de estos ejercicios son fuertemente visuales y por lo tanto accesibles también a los sectores no alfabetizados. La participación en ejercicios de DRP tiene un costo real para la comunidad, sin embargo. Como apuntó la evaluación realizada por Solidaridad Suiza 10 años después del tsunami de 2004 en Asia durante la investigación cualitativa inicial: Las reuniones formales organizadas, como discusiones de grupos focales y herramientas de diagnóstico rural participativo (DRP), no siempre arrojaron información confiable, ya que las comunidades habían sido estudiadas hasta la saciedad durante los años que siguieron al tsunami. El uso de estos métodos se limitó por lo tanto a situaciones en las que el equipo sentía que podrían añadir valor a la investigación, y se aplicó una metodología basada en entrevistas semiestructuradas “informales” más abiertas. (Ferf y Fabbri, 2014: 18-19)

Cuadro 14.2: Técnicas de DRP comunes

Técnica	Descripción	Uso potencial
<b>Calendario</b>	El grupo construye un diagrama que revela los cambios a través del tiempo – por ejemplo, en la faena agrícola, la carga de trabajo de distintos géneros o el patrón de enfermedades.	Un calendario de 24 horas construido con las mujeres podría revelar la cantidad de tiempo que invierten las mujeres en obtener artículos de socorro (por ejemplo, ayuda alimentaria) en las etapas iniciales de una crisis.  Un calendario estacional podría sacar a la luz los períodos de mayor escasez de alimentos, los mismos que la evaluación podría analizar posteriormente para determinar si coinciden con las fechas de distribución de alimentos.
<b>Línea de tiempo</b>	El grupo elabora una línea de tiempo de los acontecimientos.	Una primera línea de tiempo podría registrar los acontecimientos más importantes o los momentos de mayor inseguridad de una crisis asociada a un conflicto, mientras que una segunda línea de tiempo podría registrar las fechas de entrega de la asistencia humanitaria.
<b>Agrupamiento proporcional</b>	Se entrega al grupo una pila de 100 piedras, granos o semillas, que representan el total de una categoría dada (por ejemplo, los ingresos del hogar) y se les pide que dividan la pila en grupos para ilustrar la importancia relativa de diversos elementos dentro de esa categoría (por ejemplo, las fuentes de ingresos).	Conocer la importancia relativa de las diferentes fuentes de ingresos puede arrojar luz sobre la importancia relativa de una intervención de socorro específica, por ejemplo un programa de dinero por trabajo o de ayuda alimentaria.  Esta técnica puede usarse asimismo para identificar a los individuos o grupos más pobres de una comunidad.
<b>Jerarquización</b>	Se pide al grupo que jerarquice diferentes artículos, ya sea comparándolos unos con otros o en dos dimensiones, de acuerdo a determinados criterios.  En la jerarquización por parejas se elabora una matriz con los mismos artículos en ambos ejes y se invita al grupo a jerarquizar cada artículo en comparación con los artículos que figuran en el otro eje.	La jerarquización podría utilizarse para establecer hasta qué punto los diferentes tipos de asistencia humanitaria (por ejemplo, ayuda alimentaria, artículos no comestibles, semillas y herramientas) satisfacen las necesidades de los beneficiarios. Si se hace en dos dimensiones, podría capturar las diferentes necesidades del hogar – por ejemplo, de los adultos mayores, las mujeres, los hombres o los niños y niñas.
<b>Caminata transversal</b>	Los evaluadores recorren la aldea a pie con un grupo pequeño o con informantes clave y hacen preguntas acerca de lo que ven – por ejemplo, quién tiene acceso a los pastos de un pastizal o quiénes viven en una vivienda.	Esta técnica puede ayudar a los evaluadores a establecer los efectos diferentes de un desastre natural en distintas zonas o sectores de una comunidad, y a explorar si la respuesta humanitaria tuvo en cuenta dichas diferencias. Ofrece buenas oportunidades para la observación.
<b>Cartografía social</b>	Metodología visual para mostrar la ubicación relativa de los hogares y la distribución de los diferentes tipos de personas (por ejemplo, hombres, mujeres, adultos, niños, con tierras, sin tierras, alfabetizados y no alfabetizados) así como la estructura social y las instituciones en una zona determinada.	Puede ayudar a los evaluadores a entender cuál es la relación entre ubicación y vulnerabilidad en una comunidad. Puede ayudar a los evaluadores a enfocarse en grupos específicos para responder las preguntas referidas a la cobertura de la ayuda.
<b>Diagrama de Venn</b>	Diagrama de círculos que representan distintas categorías y se superponen, donde las categorías tienen un valor en común establecido.	Estos diagramas suelen utilizarse para mostrar las diferentes instituciones presentes en una zona – por ejemplo, organizaciones humanitarias internacionales, nacionales y locales – así como la forma en que se relacionan entre sí y su importancia relativa en la respuesta a una crisis.

11 / Diseños de evaluación para responder las preguntas de evaluación 192 - 213

12 / Muestreo 214-230

13 / Metodologías de campo 231 - 267

14 / Involucrando a la población afectada en su evaluación 268-282

15 / Acceso restringido 283 - 292

16 / Análisis 293 - 321

Para mayor información sobre el uso de metodologías de DRP, ver *Methods for Community Participation* (Metodologías de participación comunitaria; Kumar, 2002) y *Participatory Rapid Appraisal for Community Development* (Diagnóstico rápido participativo para el desarrollo de la comunidad; Theis y Grady, 1991). El enfoque de DRP se origina en el trabajo de Robert Chambers (1994). Recursos en línea disponibles en: Banco Mundial (Banco Mundial, 2007) y Agencia Canadiense para el Desarrollo Internacional (CIDA, 2005).



### Consejo útil

Los métodos de DRP parecen simples pero requieren buenas competencias de una persona que facilite. Deben ser usados con grupos lo más homogéneos posible, por ejemplo mujeres en situación de desplazamiento reciente o residentes que permanecieron en un asentamiento o ciudad después de un ataque. Puede ser útil llevar a cabo ejercicios de DRP con hombres y mujeres, o con dirigentes y no dirigentes, por separado.

El facilitador de un ejercicio de DRP deberá manejar la fina línea divisoria entre control comunitario y captura del proceso por parte de una élite (Adbullah et al., 2012).



### Ejemplo de buena práctica: DRP en una evaluación participativa en Uganda

La evaluación del programa Escuelas de Campo para Agricultores (ECA) en Uganda (Foley, 2009) hizo un uso extensivo de herramientas de DRP. Se trata de una intervención de seguridad alimentaria y medios de vida, similar a muchas intervenciones de reconstrucción en situaciones humanitarias.

Los 12 grupos del programa utilizaron las siguientes herramientas:

- Jerarquización de la riqueza, para establecer las percepciones de riqueza relativa y cómo podrían haberse reflejado en la focalización del programa;
- Calendario estacional, para establecer los ciclos productivos y los insumos requeridos y juzgar hasta qué punto habían sido considerados en la programación;
- Agrupamiento proporcional, para indagar sobre los alimentos, ingresos y gastos del hogar y sugerir qué tipos de cambios podrían esperarse de un programa de ECA efectivo;
- Jerarquización por parejas, para explorar la severidad relativa de las estrategias de manejo y articular estas ideas con los conceptos de vulnerabilidad y estacionalidad.

## 14.5 Desafíos particulares de involucrar a la población afectada en la EAH

Los desafíos de involucrar a la población afectada y sus posibles soluciones se presentan en el [Cuadro 14.3](#) a continuación.

**Cuadro 14.3:** Desafíos comunes de involucrar a la población afectada

Desafío	Posibles soluciones
Obtener acceso a la población afectada, especialmente en las crisis asociadas a conflictos donde reina la inseguridad	Encuestas telefónicas; reunirse con la población afectada en ambientes más seguros, por ejemplo en pueblos que actúan como mercados locales y/o nacionales. Ver la <a href="#">Sección 15: Acceso restringido</a> .
Barreras lingüísticas y culturales	Cerciorarse de que el equipo responsable del trabajo de campo posea las capacidades lingüísticas necesarias o tenga acceso a intérpretes;  Contratar investigadores/evaluadores locales para que se sumen al equipo.
Desarrollar metodologías robustas para trascender lo anecdótico	Prestar atención al diseño de las metodologías cualitativas, por ejemplo utilizando revisores de pares para asesorar en lo concerniente a la solidez de las metodologías;  Utilizar la <a href="#">triangulación</a> .
Las personas afectadas disponen de poco tiempo para la consulta	Estar atentos a las presiones de tiempo y diseñar sus metodologías de consulta de forma tal que no interfieran con ellas, por ejemplo consultar a la gente mientras espera en cola la distribución de la ayuda o por la tarde, al volver de sus labores.
Intento de “guardianes” de influir en la selección de los grupos con que puede reunirse el equipo de evaluación	Estar atentos a la dinámica política y de poder de las comunidades y adaptar el enfoque de las entrevistas teniendo este factor presente. Ver el consejo útil sobre “guardianes” <a href="#">más abajo</a> .
Cómo consultar con una población traumatizada	Cerciorarse de que los evaluadores posean las capacidades y la experiencia necesarias para involucrar a la población afectada teniendo en consideración el estado en que se encuentra. Ver también el <a href="#">Cuadro 14.4</a> .

11 / Diseños de evaluación para responder las preguntas de evaluación 192 - 213

12 / Muestreo 214 - 230

13 / Metodologías de campo 231 - 267

14 / Involucrando a la población afectada en su evaluación 268 - 282

15 / Acceso restringido 283 - 292

16 / Análisis 293 - 321

**Consejo útil**

Al consultar con la población afectada, estén alerta a la influencia de “guardianes” de la comunidad, por ejemplo dirigentes comunales que desean influir en la selección de las personas o grupos que debe consultar el equipo de evaluación. Es posible que deban hacer un esfuerzo extraordinario para llegar a los sectores marginados, por ejemplo, aquellos que viven en la periferia de la aldea o en la parte trasera del campamento; pero cerciórense de no ponerlos en riesgo por hablar con ellos (ver el [Cuadro 14.4](#)). Verifiquen con otros informantes clave la información referida a los diferentes grupos de la comunidad.

**Consejo útil**

Al consultar con la población afectada, consideren utilizar puntos de entrada distintos a los que emplean las organizaciones que están realizando evaluaciones para sus diagnósticos de necesidades y actividades de seguimiento permanentes, con la finalidad de evitar los sesgos y llegar a aquellos a los que nadie escucha. Por ejemplo, realicen una caminata transversal por la comunidad o el campamento para refugiados y seleccionen aleatoriamente los hogares a visitar y entrevistar (ver la [Sección 11, Diseños de evaluación para responder las preguntas de evaluación](#)).

El programa para escuchar a la población afectada impulsado por CDA describe cómo: Nosotros desde luego nos atuvimos al protocolo y nos reunimos con los jefes de aldea – y líderes comunitarios y funcionarios estatales, etc. – pero siempre pedimos autorización para reunirnos con otras personas y en ocasiones tuvimos que insistir en que nos concedieran tiempo para reunirnos con ellas (mujeres, jóvenes, grupos marginados). Le preguntamos al personal de ONG y OBC (organizaciones comunales más cercanas al llano) locales con quiénes debíamos hablar, y les pedimos a los directores de escuela, maestros, enfermeras, empresas locales, conductores de mototaxi, etc., que nos hicieran sugerencias. (Ver Anderson, Brown, Jean, 2012)

## 14.6 Consideraciones éticas de involucrar a las poblaciones afectadas

Existe una serie de consideraciones éticas que debemos tener en cuenta a la hora de planificar una evaluación (ver la [Sección 2, La decisión de realizar una evaluación](#)). La presente subsección profundiza en temas y consideraciones que son especialmente importantes a la hora de involucrar a las poblaciones afectadas.<sup>13</sup> El [Cuadro 14.4](#) contiene un resumen de algunos desafíos clave y sus posibles respuestas éticas.

Cuadro 14.4: Principales desafíos de involucrar a las poblaciones afectadas y posibles respuestas

Desafío	Respuesta ética
¿Se encuentran algunos sectores de la población en riesgo de abuso y maltrato o de violencia focalizada?	No entrevistar a las personas a no ser que sea seguro hacerlo.
¿Hay algunos grupos políticamente poderosos que dominan el proceso de consulta?	Estar atentos a la dinámica de poder y buscar formas de llegar a las personas y grupos marginados y menos poderosos.
¿Están traumatizados algunos miembros de la población afectada a raíz de sus experiencias relacionadas con la crisis?	Entrevistar a las personas con delicadeza, evitando afligir a los entrevistados, por ejemplo haciéndolos revivir una experiencia traumática. Si una pregunta tiene altas probabilidades de provocar un nuevo trauma en el entrevistado, preguntarse si la información que se desea recabar es realmente necesaria. Desde el punto de vista de la ética, es inapropiado hacer preguntas sensibles si los evaluadores no están en capacidad de remitir a las personas traumatizadas a un servicio adecuado.
¿Se ha producido una ruptura de confianza y en las relaciones entre los miembros de la población afectada a consecuencia de la crisis?	Invitar a miembros respetados de la comunidad a presentar la evaluación y el proceso de consulta.  Cerciorarse de explicar con total transparencia el propósito, las restricciones y el uso previsto de la evaluación, así como la forma en que se procesará y almacenará la información que se recoja.

Adicionalmente a las consideraciones asociadas al precepto de “no hacer daño”, que se explican en la Sección 2, La decisión de realizar una evaluación, la protección y la confidencialidad de la información, así como el consentimiento previo e informado, constituyen salvaguardas importantes que deben ser establecidas antes de llevar a cabo las entrevistas y recolectar información primaria en cualquier evaluación, pero especialmente en la EAH, donde hay aspectos de protección que considerar.

La **confidencialidad** y su relación con la **protección de información** deberían estar incluidas en el código de ética de cualquier evaluador. Esto incluye los procedimientos para garantizar la privacidad de todos los encuestados durante la fase de recolección de información. Los principios generales que subyacen a la protección y confidencialidad de la información en la evaluación son: todas las personas “son dueñas” de sus propias experiencias de vida y cualquier información que pueda ser atribuida a una persona en particular debe estar disponible para el evaluador solo si ha sido negociado previamente (Kushner, 2005: 74). En otras palabras, el nombre de un entrevistado de una población afectada no debe figurar en ningún informe de evaluación salvo que la persona en cuestión haya dado su consentimiento expreso.

Asimismo, es importante obtener el **consentimiento informado** de los miembros de la población afectada antes de entrevistarlos, especialmente si hay aspectos de protección que considerar. Esto requiere que los equipos de evaluación:

- Se cercioren de que todos los potenciales encuestados, incluidos los niños, niñas y adolescentes, entiendan perfectamente lo que entraña su participación.
- Alienten las preguntas y aclaraciones para despejar cualquier duda acerca de la evaluación antes de proceder con las entrevistas u otros ejercicios de recopilación de información.
- Programen el tiempo suficiente para que los potenciales participantes reflexionen acerca de su participación y tomen la decisión de hacerlo.
- Consideren las capacidades del entrevistador (algunos equipos de evaluación incluyen personas con capacidades básicas de asesoría por si hubiera cuestiones de protección que explorar).
- Entreguen a los entrevistadores información sobre los servicios a disposición de los entrevistados (por ejemplo, servicios sociales y de salud). Tengan presente que para todos los ejercicios de recolección de información previstos para abordar temas de violencia sexual rigen consideraciones especiales.
- Reafirmen a los participantes que tienen derecho a retirarse de la entrevista en cualquier momento.
- Consideren cuál es la mejor estrategia para garantizar que los entrevistadores cumplan con los procedimientos éticos desarrollados para la evaluación y sean coherentes en su enfoque (UNICEF 2015: 131-132, basado en el Grupo de Referencia sobre Vigilancia y Evaluación para la Protección Infantil, CP-MERG, 2012).

# 15 / Acceso restringido

## 15.1 ¿Qué es el acceso restringido?

La mayoría de las metodologías del trabajo de campo para EAH requieren que los evaluadores observen y realicen entrevistas y ejercicios con las partes interesadas. Pero, como se señaló en la [Sección 1](#), los evaluadores a menudo tienen escaso o ningún acceso a la población que ha recibido la asistencia humanitaria, o incluso a las regiones donde reside. Esto es especialmente cierto en situaciones de conflicto. Los riesgos van desde que las organizaciones humanitarias (y quienes evalúan su trabajo) se conviertan en blanco directo de los ataques (por ejemplo, si son asociadas con alguna de las partes en conflicto), pasando por verse inadvertidamente atrapadas en un conflicto en marcha, hasta poner en peligro la integridad de la población afectada por intentar obtener información de ella.

El acceso restringido<sup>14</sup> claramente constituye un problema para las personas que están ejecutando un programa humanitario así como para aquellas que lo están evaluando. Mientras que un nivel de riesgo relativamente alto puede ser aceptable como el costo que hay que pagar por proporcionar asistencia de vida o muerte, para llevar a cabo una evaluación puede ser inaceptable. La decisión de asumir riesgos debe ser siempre proporcional a los probables beneficios.

¿A qué nos referimos con “acceso restringido”? Algunos ejemplos de acceso restringido en diferentes escenarios y grados incluyen:

- Un contexto demasiado inseguro para cualquier tipo de evaluación formal, por ejemplo en algunas regiones de Somalia y Siria.
- Mientras que los evaluadores locales pueden tener un mayor acceso porque son más aceptados y menos visibles que los extranjeros, por ejemplo en la mayor parte de Afganistán y durante el conflicto de Darfur; en otros casos, como el conflicto de los Balcanes o las Islas Molucas del Norte, el personal nacional puede estar bajo mayor amenaza que los expatriados.
- Es posible que el equipo de evaluación en pleno pueda llegar a algunos sectores de la población afectada, por ejemplo a las personas desplazadas internamente que permanecen en o cerca de centros poblados, pero no a las zonas rurales, como sucedió durante el conflicto de Darfur.
- El acceso de evaluadores internacionales a un país o a ciertas regiones puede estar sujeto a restricciones administrativas, por ejemplo la necesidad de visas, permisos de viaje o certificados de no objeción.

El acceso restringido representa un desafío asimismo para el seguimiento de las intervenciones humanitarias y puede limitar el seguimiento de la información en entornos inseguros.

Esta sección se enfoca en cómo decidir si proceder o no con una evaluación en circunstancias en que el acceso es restringido. Se examinan distintos enfoques de evaluación en estas situaciones y cuándo es posible que una evaluación tenga que llevarse a cabo a distancia. Así como la gestión a distancia se encuentra a la vanguardia en el campo de la asistencia humanitaria, la evaluación a distancia se encuentra también de muchas formas a la vanguardia en la EAH. Diversas organizaciones y evaluadores vienen experimentando con diferentes enfoques, algunos de los cuales, los más prometedores, se presentan aquí.<sup>15</sup> La [Sección 15.5](#) examina opciones para asegurar la credibilidad de una evaluación cuando el acceso es restringido y una parte, sino toda, tiene que hacerse a distancia.

## 15.2 Importancia del análisis de evaluabilidad

Cuando el acceso es restringido, es vital realizar un análisis de factibilidad o evaluabilidad antes de empezar una EAH. [La Sección 2: La decisión de realizar una evaluación](#), explica el concepto básico de análisis de evaluabilidad como un medio para contar con una base clara a partir de la cual proceder, y para hacer una evaluación tan útil como sea posible. Si el acceso es restringido, el análisis deberá enfocarse en si es factible o no evaluar una actividad o un programa dado con credibilidad, y si vale la pena hacerlo.



### **Ejemplo de buena práctica: La Unidad de Evaluación de Médicos Sin Fronteras (MSF)**

Apenas recibimos una solicitud de evaluación, la Unidad de Evaluación de MSF se aboca a estudiar detenidamente los términos de referencia y posibles opciones de diseño. La unidad considera qué ajustes se pueden hacer a la “norma dorada” de diseños que se ajusten al contexto específico de una operación. En situaciones de acceso restringido, por ejemplo operaciones transfronterizas en Siria, la dificultad más importante reside en obtener información primaria sobre los resultados así como retroalimentación directa de los beneficiarios. En consecuencia, el diseño de evaluación se enfocará en el proceso y la actividad más que en los resultados y tendrá que apoyarse más en información secundaria y percepciones. Estas opciones y limitaciones son compartidas con toda claridad con la entidad que encarga la evaluación, de modo tal que puedan determinar si una evaluación se ajustaría a sus necesidades de aprendizaje y rendición de cuentas o si deberían contemplar otras opciones.

Fuente: Sabine Kampmueller, Unidad de Evaluación de MSF, comunicación personal (septiembre de 2015)



**Las preguntas centrales que se deben plantear para responder si vale la pena realizar una evaluación incluyen:**

1. ¿Cuáles son los principales riesgos que enfrenta la evaluación en este contexto? Estos pueden incluir:
  - a. riesgos operativos, tales como la seguridad personal de los evaluadores y la posible suspensión del programa si se desvían recursos para financiar la evaluación;
  - b. riesgos financieros, si el hecho de trabajar en un entorno inseguro acarrea costos adicionales; y
  - c. riesgos de protección a los que puede exponerse a la población afectada si participa en la evaluación.
2. ¿Cuáles son las implicaciones de los riesgos mencionados para el acceso de los evaluadores a las poblaciones afectadas?
3. ¿Qué información secundaria y otros tipos de información estarán disponibles si los evaluadores no pueden obtener acceso a la población afectada?
4. ¿Qué otras opciones existen para que los evaluadores obtengan acceso a la población afectada (ver la [Sección 15.3 sobre metodologías de evaluación a distancia](#))?
5. ¿Cómo afectará este hecho la credibilidad de la evaluación?
6. ¿Cuáles son las alternativas posibles a una evaluación, por ejemplo un taller de aprendizaje reflexivo con el personal, aprendizaje de pares entre organizaciones o una actividad evaluativa más limitada?

El análisis de evaluabilidad puede ser particularmente importante cuando las organizaciones sienten que se encuentran bajo la presión de donantes para cumplir la obligación contractual de evaluar sistemáticamente si una evaluación es viable.



**Ejemplos de buenas prácticas: Análisis de evaluabilidad**

En 2009, el Comité Permanente Interagencial (IASC) mandó hacer un análisis de evaluabilidad (Cosgrave, 2010) de una propuesta de evaluación de intervenciones humanitarias en el centro y sur de Somalia.

Además de incidir en los aspectos que suelen formar parte de un análisis de evaluabilidad, por ejemplo la definición de los objetivos y los alcances de la evaluación, la pregunta clave era cómo llevar a cabo una evaluación cuando incluso el seguimiento más sencillo era muy difícil. El elemento principal del análisis de evaluabilidad fue un comentario, sustentado en entrevistas con informantes clave y una revisión documental, sobre un borrador de los TdR de la evaluación que se remontaba a 2008. La evaluación se llevó a cabo con éxito en 2011 (Polastro et al., 2011).





En 2014, Global Affairs Canada, el ministerio de asuntos globales de Canadá, realizó un “análisis de escenarios y riesgos” antes de emprender la evaluación de su programa nacional en Sudán del Sur. La pregunta central fue cómo afectaría el conflicto en marcha en Sudán del Sur a la fase inicial y a la recolección de información. Una pregunta subsidiaria tenía que ver con los planes de contingencia necesarios. Un análisis sistemático de potenciales escenarios exploró la situación en ese momento y estimó una serie de resultados posibles. Acto seguido se elaboró un registro de los riesgos, distinguiendo entre riesgos operativos, financieros, para la calidad del informe y para la reputación de la institución. Sobre la base de este análisis se decidió que la evaluación debía proceder y se adoptaron diversas medidas para mitigar algunos de los riesgos.

## 15.3 Opciones para enfrentar el acceso restringido

Las organizaciones y evaluadores han experimentado con una serie de opciones creativas para hacer frente al acceso restringido a las poblaciones afectadas. Algunas de las más comunes se muestran en el [Cuadro 15.1](#).

Cuando el acceso es restringido, es posible que los evaluadores tengan que apoyarse en gran medida en información secundaria, la misma que sin embargo con frecuencia es precaria en dichos contextos.



### **Definición: Crowd-sourcing (colaboración abierta distribuida)**

En el método de *crowd-sourcing* se utiliza un gran número de voluntarios, ya sea para recolectar información o para analizar la información de imágenes digitales, generalmente por medios indirectos. Este tipo de información se denomina “datos obtenidos por crowd-sourcing”.

## Ejemplo de crowd-sourcing para hacer seguimiento: Imágenes satelitales analizadas por voluntarios

En 2011, ACNUR reclutó la ayuda de voluntarios para etiquetar tres tipos de alojamiento informal, con miras a obtener una estimación de la población desplazada internamente en el Corredor de Afgooye en Somalia. Los voluntarios procesaron 3,909 imágenes satelitales en solo cinco días y añadieron más de 250,000 etiquetas.

Fuente: <http://irevolution.net/2011/11/09/crowdsourcing-unhcr-somalia-latest-results>

Cuadro 15.1: Opciones para enfrentar el acceso restringido

Técnica	Cómo usar esta metodología	Posibles escollos
<p>Utilizar investigadores y/o evaluadores locales para llevar a cabo entrevistas con la población afectada (Norman, 2012: 30-35).</p> <p><u>(Ver ejemplo de buena práctica en la pág. 289.)</u></p>	<p>Planificar un taller de capacitación desde el principio (Sida, 2013), así como un taller de análisis una vez concluido el trabajo de campo.</p> <p>De ser apropiado, pedir a los investigadores y/o evaluadores locales que graben sus entrevistas para poder verificarlas.</p> <div style="border: 1px solid gray; padding: 5px; margin-top: 10px;">  <p><b>Consejo útil</b> Asegúrense de programar el tiempo suficiente para estos talleres, cuya duración dependerá de las capacidades y experiencia de los investigadores o evaluadores locales.</p> </div>	<p>Factible únicamente si no se pone en riesgo la integridad de los investigadores y/o evaluadores.</p> <p>En conflictos de larga duración puede ser difícil encontrar investigadores y/o evaluadores que no sean percibidos como vinculados a una de las partes.</p>
<p>Realizar encuestas en línea, por teléfono y/o por mensajería de texto (ver ejemplo de buena práctica en la pág. 290, así como la Sección 13, Metodologías de campo).</p>	<p>Esta opción puede utilizarse para encuestas relativamente cortas y directas, por ejemplo para averiguar cuándo recibió asistencia la población, qué recibió y cuánto. Las encuestas telefónicas pueden usarse también con el personal de las oficinas de campo.</p> <p>Las líneas directas pueden ser una oportunidad para que la población afectada plantee temas de interés, especialmente si fueron instaladas durante la ejecución del programa (Walden, 2013: 3).</p>	<p>Las encuestas en línea y por mensajería de texto están sujetas al sesgo de la autoselección y deben ser interpretadas con cautela.</p> <p>Las encuestas telefónicas también pueden entrañar sesgos, por ejemplo al ser accesibles para aquellas personas que poseen un teléfono móvil solamente.</p> <p>En un ambiente altamente politizado, es posible que la población local no confíe en y/o se muestre reticente a utilizar una línea directa.</p>





Cuadro 15.1: Opciones para enfrentar el acceso restringido

Técnica	Cómo usar esta metodología	Posibles escollos
<p>Entrevistar a miembros de la población afectada en zonas accesibles.</p> <p>(Ver ejemplo de <u>buena práctica</u> en la pág. 290, así como la <u>Sección 13, Metodologías de campo.</u>)</p>	<p>Averiguar si miembros de la población afectada se desplazan con regularidad a otras zonas más accesibles, por ejemplo a pueblos en los que hay mercados más seguros, y organizar entrevistas y discusiones de grupos focales con estas personas.</p> <p>Pedir a los miembros de la población afectada que viajen a zonas que sean accesibles para el equipo de evaluación, siempre que no comprometa su seguridad personal.</p>	<p>Puede introducir sesgos, por ejemplo si viajar se considera seguro para los hombres, pero no para las mujeres.</p> <p>Es difícil triangular los hallazgos si el equipo de evaluación solo puede reunirse con ciertos grupos.</p>
<p>Observación a distancia (Norman, 2012: 45-48)</p> <p>(Ver la <u>Sección 13: Metodologías de campo.</u>)</p>	<p>Pueden usarse imágenes satelitales para inspeccionar la infraestructura (construida en el marco de programas de dinero por trabajo, por ejemplo), o para revisar los patrones de asentamiento.</p> <p>Se puede pedir a informantes clave y miembros de la población afectada que tomen videos y fotografías, utilizando cámaras con GPS incorporado.</p> <p>Si los programas están realizando un seguimiento a distancia, por ejemplo al haber proporcionado tablets, smartphones o cámaras al personal, éste puede aprovecharse para recolectar información.</p>	<p>Si bien estas técnicas pueden ser apropiadas para observar situaciones físicas e infraestructura, posiblemente revelen poco acerca del uso que se les está dando a las mismas, por ejemplo, quién tiene acceso a obras de infraestructura específicas, como puntos de agua.</p>
<p>Datos obtenidos por crowd-sourcing, por ejemplo por Twitter y Facebook (Sida, 2014)</p>	<p>Se podría utilizar información obtenida por crowd-sourcing para analizar cuán difundido está el uso de ciertas instalaciones, rastreando el uso de los teléfonos móviles.</p> <p>Verificar si se ha hecho un seguimiento a distancia a través de medios sociales como Facebook y Twitter, y si se podría utilizar la información en la evaluación y cómo (ver <u>ejemplo de buena práctica</u> en la pág. 257).</p>	<p>Los encuestados se seleccionan a sí mismos, con lo cual se está introduciendo un sesgo; por ejemplo, los jóvenes utilizan los medios sociales más que los adultos mayores.</p> <p>Es posible que las partes en el conflicto hayan realizado campañas en los medios sociales.</p>

Si el programa ha desarrollado sistemas de seguimiento a distancia durante la fase de ejecución, quizás sea posible capitalizar dichos sistemas. Por ejemplo, si se han instalado centros de llamadas (call centres) o líneas directas, estos podrían ser aprovechados en la evaluación para recolectar información. Ver ejemplo de buena práctica en la pág. 290.

**Tener presente**

Como sucede en todas las evaluaciones, el proceso de recolección de información no debe suponer un riesgo para la integridad de las personas.



### **Ejemplo de buena práctica: Evaluación de la cooperación de la FAO en Somalia 2007 – 2012**

Como parte de esta evaluación, y reconociendo el acceso restringido que tenía el equipo a muchas regiones geográficas de Somalia, se encargó un estudio en profundidad del programa de dinero por trabajo (DPT) ejecutado por la FAO. El trabajo de campo se llevó a cabo en el transcurso de un mes. Se designó a una investigadora-evaluadora internacional (una antropóloga) para encabezar el estudio. A pesar de carecer de acceso a las ubicaciones del programa por razones de seguridad, la evaluadora contrató a cinco investigadores locales cuidadosamente seleccionados mientras que ella trabajaba desde Dollow, bastante cerca de las ubicaciones del programa de DPT. Se programó un período de cinco días para capacitar a los investigadores locales, que utilizaron metodologías cualitativas de DRP para recolectar la información. Se aplicó una técnica de bola de nieve (ver la Sección 12, Muestreo), a través de la cual se recolectó información de manera iterativa durante un mes. Los investigadores locales visitaron cada aldea por espacio de tres días, retornando a Dollow para informar sobre el trabajo realizado y analizar sus hallazgos con la jefa del equipo. Fue una oportunidad para examinar temas imprevistos, a menudo generando preguntas adicionales para ser planteadas en la siguiente visita de campo. Las técnicas y herramientas de DRP generaron materiales fácilmente analizables e indispensables para la investigación iterativa y el proceso de capacitación. Se usó la triangulación con base en registros fotográficos y grabaciones de audio realizadas por los investigadores locales. Utilizando la información del centro de llamadas que instaló la FAO para fines de seguimiento durante la ejecución del programa, se contrató a otro consultor somalí para que llevara a cabo encuestas telefónicas, las mismas que también se usaron para triangular la información recogida en el campo, contrastándola con las respuestas de beneficiarios del programa de DPT de otras regiones.

Adicionalmente al estudio del programa de DPT, el equipo de evaluación invitó asimismo a algunos “intermediarios” (que podían viajar), por ejemplo a miembros de las asociaciones de agricultores y ganaderos, a acudir a Mogadishu para reunirse con el equipo en un espacio seguro.

Fuentes: Buchanan-Smith et al. (2013); Tessitore (2013)



### **Consejo útil**

Consideren si ciertos evaluadores podrían estar expuestos a menos peligros en razón de su nacionalidad o identidad étnica, por ejemplo. No olviden, sin embargo, que tienen la obligación de velar por la seguridad de todos los evaluadores.



### **Ejemplo de buena práctica en entrevistas telefónicas: Evaluación de Oxfam GB de la respuesta a la sequía de 2011– 2012 en Somalia**

Para su programa de distribución de dinero en efectivo, Oxfam GB recolectó los números de teléfono móvil de los beneficiarios participantes (la ETR sugiere que entre el 10 y el 15 por ciento de los beneficiarios inscribieron su número de teléfono) y, durante la ETR realizada en septiembre de 2011, llamaron a un pequeño número de personas para que comentaran el programa. Esto arrojó algunos resultados notables. Por ejemplo, de los 12 números contactados, cinco respondieron (uno de ellos incluso devolvió la llamada perdida). El propósito de estas conversaciones era evaluar cuánto sabían los beneficiarios de los criterios de selección, su impresión del proceso, su percepción del proyecto y si sabían cómo presentar un reclamo. La retroalimentación que se recibió a través de esas llamadas arrojó mucha luz sobre diversos aspectos. Por ejemplo, todos los encuestados aseguraron que entendían los criterios de selección y la mayoría indicó que sabía cómo contactar a un funcionario de la organización en caso necesario, pero ninguno sabía cuándo, cómo y cuánto dinero iban a recibir.

Fuente: Featherstone (2012: 13-14)



### **Ejemplo de buena práctica: Llevando a miembros de la población afectada a zonas más accesibles**

Al evaluar un programa con personas desplazadas internamente (PDI) en la República Democrática del Congo (RDC), para Groupe URD, el equipo de evaluación, era importante hablar directamente con la población afectada. Sin embargo, habrían tenido que transitar por territorio rebelde para obtener acceso a los asentamientos. Aunque los extranjeros corrían el riesgo de ser secuestrados, los desplazados internos varones podían atravesar la zona con relativa seguridad. Las contrapartes locales habían aprovechado este hecho anteriormente para encomendarles la distribución de la asistencia. Por consiguiente, Groupe URD desarrolló criterios específicos para ayudar a la población desplazada internamente a seleccionar una gama de representantes (por ejemplo, agricultores, jefes religiosos o tradicionales, etc.). Así, se seleccionó a cuatro o cinco representantes por asentamiento y se les pidió desplazarse hasta una aldea segura para ser entrevistados. Los desplazados internos también tenían la opción de identificar a alguna persona que ya estuviera en la aldea de destino y consideraran capacitada para hablar en su nombre.

Para que el hecho de desplazarse valiera la pena, las contrapartes locales ofrecieron a los representantes suministros que pudieran llevar consigo de regreso a sus comunidades.

La desventaja de este tipo de enfoque es que los evaluadores no pueden controlar a quién escogerá la comunidad. Además, solo pueden ir aquellos que puedan pasar por la zona insegura (por ejemplo, para las representantes femeninas no es seguro viajar en ciertas circunstancias).

Fuente: Bonaventure Sokpoh, Groupe URD, comunicación personal, 2015



### Ejemplo de buena práctica: Evaluación a distancia basada en entrevistas con el personal – Médicos Sin Fronteras en Pakistán

En 2009, MSF inició una evaluación intersectorial de sus proyectos manejados a distancia en todo el mundo. Por razones de seguridad, en Pakistán el evaluador tenía un acceso limitado al programa o a la población afectada. Para poder formarse una imagen más completa del programa, el evaluador se puso en contacto con personal presente y pasado de MSF. El personal internacional fue fácilmente ubicado (y entrevistado en reuniones, ya sea personales o telefónicas), pero con el personal nacional no sucedió lo mismo. El evaluador utilizó la técnica de la bola de nieve para identificar al personal con el que sería provechoso hablar y obtener información de contacto. Fue un proceso largo y tedioso pero valió la pena, ya que amplió considerablemente el tamaño de la muestra. Antes de la misión de campo, los evaluadores verificaron los indicios, pero gran parte del trabajo se hizo en el país desde la capital y desde una oficina de coordinación de proyectos (en la ciudad de Peshawar). Fue importante viajar a Peshawar, para estar lo más cerca posible del lugar del proyecto, buscar candidaturas apropiadas para las entrevistas y entender mejor el contexto.

Fuente: Mzia Turashvili, Unidad de Evaluación de MSF, comunicación personal, noviembre de 2015

## 15.4 Credibilidad de la evaluación a distancia

Si los evaluadores han tenido acceso limitado a la población afectada, esto genera la pregunta de cómo garantizar la credibilidad de los hallazgos.

Consideren hacer lo siguiente:

1. Describan explícitamente en el informe de evaluación las limitaciones y restricciones que enfrentó el equipo de evaluación y de qué forma estas afectaron a los hallazgos de la evaluación. Indiquen claramente a qué grupos poblacionales o áreas geográficas se refieren los hallazgos. Tengan cuidado en su análisis de no generalizar, haciendo extensivos sus hallazgos a grupos o áreas que el equipo no hubiera podido visitar.
2. La triangulación siempre es importante en la EAH. Cuando el acceso es restringido, la triangulación cobra aún mayor importancia. Por ejemplo, si dependen de equipos de evaluación locales que han estado trabajando por su cuenta, comparen la información recolectada por diferentes evaluadores en un mismo lugar, con el fin de identificar posibles sesgos. Triangulen la información recogida de la población afectada por medio de encuestas a distancia, por ejemplo encuestas en línea o telefónicas, con información de informantes clave de la zona y del personal de la organización.



3. Cuando sientan que sus hallazgos no pueden ser concluyentes debido a las restricciones enfrentadas, consideren convertir los hallazgos dudosos en hipótesis a ser validadas por la organización después de la evaluación. Quizás deseen recomendar estudios posteriores y recolección de información adicional con respecto a algunos temas en particular.

#### Consejo útil

Si obtienen información contradictoria de diferentes fuentes y no la pueden conciliar, y no han podido visitar a la población o zona en cuestión, estén preparados para abstenerse de usar esa información en su análisis, o dejen constancia expresa en su informe de que no han podido conciliar dos (o más) versiones diferentes.

## 15.5 Otras alternativas a la evaluación a distancia

- **Seguimiento a distancia:** Consideren utilizar un sistema de seguimiento a distancia si se llega a la conclusión de que la evaluación a distancia no es factible. El seguimiento a distancia debería ser más fácil que la evaluación si se enfoca principalmente en los insumos y productos, antes que en los efectos y el impacto más amplio. También vale la pena considerar el seguimiento por terceros, por ejemplo si se encarga a una ONG hacer seguimiento del trabajo de otra.
- **Aprendizaje de pares:** Las organizaciones que trabajan en una zona relativamente inaccesible podrían ser invitadas a juntarse para compartir aprendizajes y experiencia. Esto también podría hacerse para estimular la reflexión y el aprendizaje. En 2013 y 2014, ALNAP organizó, conjuntamente con el Comité de Emergencia para Desastres (DEC), talleres de aprendizaje de pares para el personal de la organización que estaba trabajando en evaluación, seguimiento y evaluación, rendición de cuentas y aprendizaje en relación con la respuesta a la crisis siria. Fue una excelente oportunidad para buscar asesoría y compartir aprendizajes (Sida, 2013, 2014).

En algunos casos pueden considerar aconsejar que el trabajo de seguimiento a distancia sea mejorado antes de proceder con una evaluación. Existe un número significativo de recursos y enseñanzas sobre seguimiento a distancia, entre ellos los siguientes:

- De Somalia, ver Polastro et al. (2011: 29);
- De la respuesta a Siria, ver la revisión de aprendizajes de 2014 de Acción contra el Hambre (ACF) (2014: 54-55);
- De un proyecto de investigación financiado por el Fondo de Innovación Humanitaria (HIF), ver el informe de Tearfund sobre prácticas de seguimiento y rendición de cuentas para proyectos manejados a distancia, implementados en entornos operativos volátiles (Norman, 2012).

# 16 / Análisis

Esta sección incide en las metodologías utilizadas para transformar la información recogida en los hallazgos que conforman la base de las conclusiones y recomendaciones de la evaluación.

La sección contiene lo siguiente:

- Análisis requerido para responder las preguntas de evaluación;
- Análisis requerido para responder las preguntas causales;
- Análisis cualitativo de información primaria, por ejemplo apuntes de entrevistas;
- Análisis cualitativo de información secundaria, por ejemplo informes de avances;
- Análisis estadístico de información primaria, por ejemplo información de encuestas;
- Análisis numérico de información secundaria, por ejemplo de registros de distribución de la asistencia;
- De los hallazgos a las conclusiones y recomendaciones.



## **Definición: Información primaria**

Información primaria es la información recolectada para los fines de la evaluación.

Las entrevistas y encuestas que se realizan específicamente para la evaluación constituyen ejemplos de información primaria.



## **Definición: Información secundaria**

Información secundaria es la información recogida para otros fines pero que es usada por la evaluación.

La información secundaria incluye: documentos de política de la organización, informes de avance, otras evaluaciones y literatura científica relevante. Ver la [Sección 11: Diseños de evaluación para responder las preguntas de evaluación para mayores detalles sobre documentación secundaria](#).

Tanto la información secundaria como la información primaria suministran evidencia para una evaluación.



## **Definición: Evidencia**

Evidencia es el cuerpo de hechos o información disponible que indica si una proposición o creencia es verdadera.



Ver la publicación de ALNAP para mayor información sobre la calidad de la evidencia (Knox-Clarke y Darcy, 2014).

#### Consejo útil

Los cuadros de evidencia facilitan la redacción del informe, al reunir toda la evidencia relacionada con una pregunta o un tema en particular en un solo lugar. Registren en un cuadro de evidencia toda la evidencia y hallazgos emergentes recolectados en entrevistas y documentos.

Se trata de un cuadro en que el equipo de evaluación registra las piezas de información vinculadas a la pregunta o al tema de evaluación, y que sirve de precodificación de la información. Contribuye a sacar a la luz qué aspectos cuentan con un soporte de evidencia fuerte y cuáles no.

Al trabajar en los hallazgos y las conclusiones y recomendaciones, el equipo de evaluación puede ver qué preguntas o temas requieren una base de evidencia más contundente.

## 16.1 *Big-n* o *small-n*

Los enfoques básicos para analizar la información *big-n* y *small-n* son diferentes. La información cuantitativa o *big-n* suele ser analizada estadísticamente, mientras que la información cualitativa o *small-n* suele ser analizada codificando la información (ver la definición de codificación más abajo). Como se señaló en la [Sección 11](#), estas dos categorías no son herméticas y puede haber algunas superposiciones. Algunos diseños están ligados a metodologías de análisis particulares.

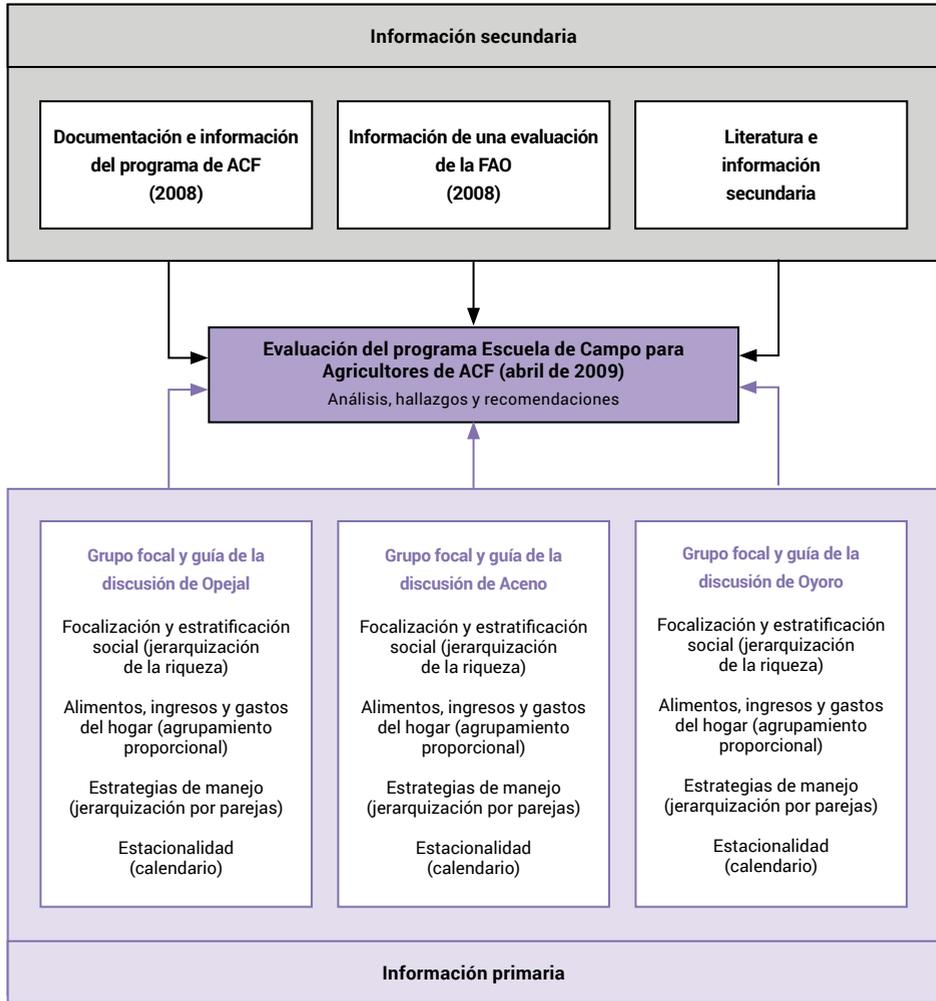
Una diferencia clave entre las metodologías *big-n* y *small-n* es la metodología utilizada para garantizar la exactitud y la fiabilidad. En las metodologías *big-n*, la exactitud y la fiabilidad se logran aplicando una metodología en particular de una forma en concreto. En las metodologías *small-n*, la exactitud y la fiabilidad se logran en parte a través de la metodología misma y en parte a través de la triangulación (introducida en la [Sección 13, Metodologías de campo](#)).

En la triangulación se encuentra implícita la idea de que las metodologías mixtas y cualitativas se apoyan mutuamente para respaldar un hallazgo y una conclusión en particular.

Los hallazgos de cualquier evaluación deben sustentarse en múltiples fuentes (ver la [Sección 10, Métodos documentales](#) y la [Sección 13: Metodologías de campo](#)). Las conclusiones, a su vez, pueden sustentarse en múltiples hallazgos. El principio central es que la evidencia debe ser triangulada, para contribuir a garantizar su calidad.

El siguiente ejemplo de una evaluación participativa en Uganda (Foley, 2009: 49) demuestra cómo se recurrió a diferentes fuentes de evidencia para contribuir a la evaluación.

Figura 16.1: Diagrama de una metodología de evaluación



11 / Diseños de evaluación para responder las preguntas de evaluación 192 - 213

12 / Muestreo 214- 230

13 / Metodologías de campo 231 - 267

14 / Involucrando a la población afectada en su evaluación 268- 282

15 / Acceso restringido 283 - 292

16 / Análisis 293 - 321

## 16.2 Análisis de evidencia para responder las preguntas normativas

Una forma de responder las preguntas normativas – por ejemplo, ¿hasta qué punto cumplió el programa de alojamiento con las normas mínimas del Proyecto Esfera? – es tratar de encontrar evidencia triangulada del cumplimiento o incumplimiento de las normas aludidas.

Si bien este enfoque puede ser apropiado para una evaluación muy básica de una intervención sencilla, puede llevar a la sugerencia de que cualquier conclusión negativa en relación con el cumplimiento se basa en anécdotas o sesgos, antes que en evidencia rigurosa.

Un enfoque más riguroso consiste en dividir la norma pertinente o el criterio aplicable en elementos individuales. (Ver la Sección 5, Enmarcando su evaluación, para ejemplos de normas o criterios que podrían ser aplicables.) Por ejemplo, para cumplir con las normas mínimas del Proyecto Esfera sobre alojamiento, es necesario cumplir primero con las seis normas mínimas esenciales del Proyecto Esfera (Proyecto Esfera, 2011: 49):<sup>16</sup>

- Respuesta humanitaria centrada en las personas
- Coordinación y colaboración
- Evaluación
- Diseño y respuesta
- Desempeño, transparencia y aprendizaje
- Desempeño de los trabajadores humanitarios

Y luego con las cinco normas mínimas detalladas del Proyecto Esfera en materia de alojamiento (Proyecto Esfera, 2011: 239):

- Planificación estratégica
- Planificación de asentamientos humanos
- Espacio vital cubierto
- Construcción
- Impacto ambiental

Si examinamos hasta qué punto una intervención ha cumplido con cada uno de estos criterios, podremos responder la pregunta global. Las normas mínimas del Proyecto Esfera incluyen indicadores que pueden usarse con algunas normas directamente (si se estima que los indicadores son apropiados para el contexto en particular). Otras normas pueden ser respondidas únicamente por medio de un juicio evaluativo, y en esos casos pueden utilizarse las técnicas descritas en la Sección 16.

## 16.3 Análisis de evidencia para preguntas evaluativas

En las preguntas evaluativas (ver la [Sección 6, Selección de las preguntas de evaluación](#)) los evaluadores deben emitir un juicio de valor. Este tipo de pregunta puede suponer un problema para los evaluadores, ya que dichos juicios pueden ser criticados por ser muy subjetivos. Esto es un problema en todos los tipos de evaluación, independientemente del diseño o las metodologías. Por ejemplo, una encuesta puede establecer que algún factor crítico ha subido en un 20 por ciento a consecuencia de una intervención, pero el evaluador tiene que emitir un juicio, es decir, debe establecer si dicha subida representa un buen o mal desempeño en las circunstancias dadas.

Una forma de afrontar las preguntas evaluativas es dividir las con el objeto de separar los elementos descriptivos, normativos y otros, del elemento puramente evaluativo. Otro enfoque reside en adoptar un enfoque riguroso y transparente para responder las preguntas evaluativas.



### **Definición: Razonamiento evaluativo**

Razonamiento evaluativo es el proceso analítico mediante el cual los evaluadores responden a las preguntas evaluativas.

El razonamiento evaluativo sintetiza la información sobre la calidad y el valor, combinando:

- Evidencia sobre el desempeño en una dimensión en particular, así como una interpretación de la misma en función de las definiciones vigentes de “buena calidad”, con el fin de generar una calificación del desempeño en la referida dimensión;
- Calificaciones del desempeño en varias dimensiones con el fin de proveer una conclusión general acerca del nivel de desempeño en general (Davidson, 2014: 1).

Ver ejemplos del uso del razonamiento evaluativo por medio de la rúbrica evaluativa según el planteamiento de Davidson (2014) más abajo. Oakden afirma que las rúbricas “ofrecen un procedimiento para explicitar los juicios en una evaluación y se usan para juzgar la calidad, el valor o la importancia del servicio prestado” (2013: 5).



### **Definición: Rúbrica evaluativa**

Una rúbrica evaluativa es un cuadro que describe cómo debería mostrarse la evidencia a distintos niveles de ejecución, en relación con algún criterio de interés o para la intervención en su conjunto.

Las rúbricas evaluativas no reciben una puntuación numérica como las rúbricas que se utilizan para el análisis documental, sino que son calificadas por medio de indicadores de calidad como “excelente” o “deficiente”. La definición de un desempeño ideal puede extraerse de normas como las normas mínimas del Proyecto Esfera (Proyecto Esfera, 2011) o bien de los documentos del proyecto o de los documentos de política de la organización.

Las rúbricas evaluativas contienen dos elementos:

- Descriptores de desempeño – deficiente, adecuado, bueno, excelente, etc.;
- Criterios que establecen cómo debe ser una “buena” ejecución. En el modelo de Davidson, esto se hace en forma de un listado de criterios bajo cada uno de los descriptores de desempeño para una dimensión en particular.

Desempeño	Excelente	Muy bueno	Bueno	Adecuado	Deficiente
<b>Descriptores de desempeño</b>	Claro ejemplo de desempeño ejemplar o práctica idónea en la distribución de alimentos; ninguna debilidad.	Desempeño muy bueno o excelente en todo sentido. Fuerte en general pero no ejemplar. Ninguna debilidad de consecuencias reales.	Desempeño razonablemente bueno. Puede haber algunas debilidades menores pero nada grave.	Desempeño razonable. Algunas debilidades graves pero no fatales.	Clara evidencia de funcionamiento insatisfactorio; graves debilidades a todos los niveles en aspectos cruciales.
<b>Listado de características de este desempeño</b>					

Fuente: Basado en Davidson (2004: 137)



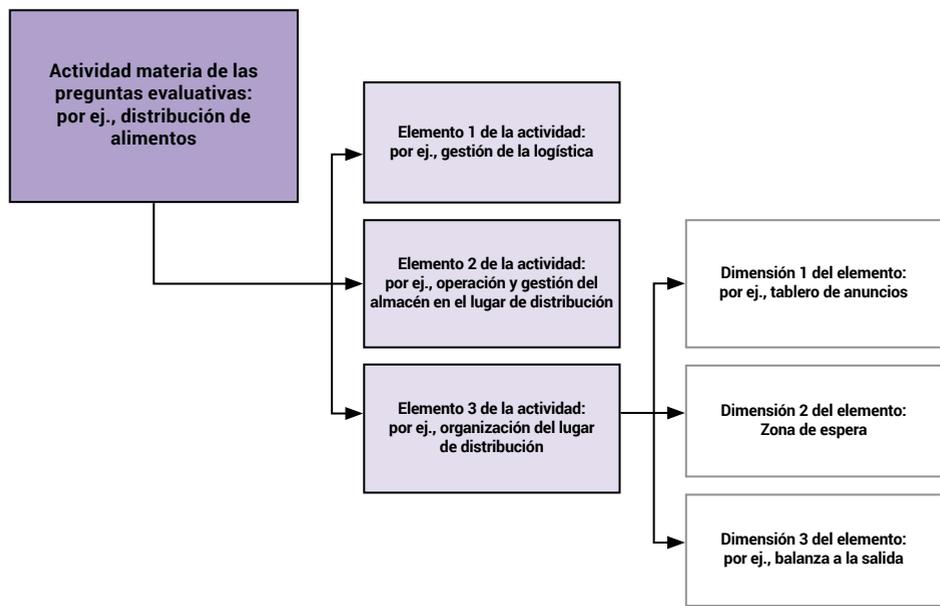
### Consejo útil

Desarrollen la rúbrica evaluativa antes de su recolección principal de información. Definan lo que a su juicio constituyen las características de los diferentes niveles de desempeño antes de iniciar su recolección de información principal. No reinventen la rueda. Revisen los modelos o normas conceptuales del sector para elaborar los criterios para su rúbrica (ver la [Sección 5, Enmarcando su evaluación](#)).

Davidson utiliza un listado de las características de cada descriptor de desempeño para cada elemento de la actividad, como en Oakden (2013: 6-7). En el ejemplo que nos ocupa, el listado de las características de un desempeño excelente en “organización de un lugar de distribución de alimentos” (un elemento de la actividad de distribución de alimentos) podría incluir las siguientes dimensiones:

- Tablero de anuncios ilustrado mostrando el volumen de las raciones para cada grupo en los idiomas pertinentes;
- Zona de espera bajo la sombra con control del acceso y acceso a agua e inodoros;
- Suficientes balanzas de gran capacidad para que los receptores puedan corroborar sus raciones – con peso de prueba;
- Clara señalización para dirigir a las personas a los pasillos adecuados;
- Pasillos limpios y secos para hacer cola;
- Cucharas en buen estado y claramente señalizadas;
- Medidas de control de colas para prevenir los aplastamientos.

En esencia, se trata de desglosar una actividad relativamente compleja en elementos y dimensiones individuales, para una evaluación más transparente de la calidad de la actividad (determinada por los criterios que figuran en [la rúbrica](#)). Por ejemplo:



La rúbrica que mostramos a continuación revela algunas de las dimensiones de la organización del elemento “lugar de distribución de alimentos” de la actividad “distribución de alimentos”:

Actividad	Distribución de alimentos				
Elemento	Organización del lugar de distribución				
Dimensión	Excelente	Muy bueno	Bueno	Adecuado	Deficiente
Tablero de anuncios	Tablero de anuncios ilustrado mostrando el volumen de las raciones para cada grupo en los idiomas pertinentes	Tablero de anuncios sin gráficos mostrando el volumen de las raciones para cada grupo en los idiomas pertinentes	Tablero de anuncios mostrando el volumen de las raciones para cada grupo pero faltan algunos idiomas clave	Tablero de anuncios mostrando el volumen general de las raciones pero falta información sobre los grupos o faltan idiomas	Tablero de anuncios inexistente
Zona de espera	Zona de espera bajo la sombra con control del acceso y acceso a agua e inodoros	Zona de espera bajo la sombra con control del acceso y acceso a inodoros	Zona de espera bajo la sombra con control del acceso pero sin inodoros	Zona de espera con control del acceso	Zona de espera formal inexistente
Balanzas a la salida	Suficientes balanzas de gran capacidad para que los receptores puedan corroborar sus raciones – con peso de prueba	Suficientes balanzas de gran capacidad para poder corroborar las raciones – pero sin peso de prueba	Balanzas de gran capacidad para poder corroborar las raciones pero se forman colas	Balanzas a la salida pero no del tamaño suficiente para pesar la ración total	Balanzas a la salida inexistentes

Criterios tomados de ACNUR (1997); Jaspers y Young (1995); Proyecto Esfera (2011)

Las rúbricas evaluativas pueden utilizar criterios de desempeño tanto cuantitativos como cualitativos. En el siguiente ejemplo sobre la calificación obtenida por la afirmación “los distribuidores trataron respetuosamente a los receptores” se utilizan las puntuaciones obtenidas en una encuesta para establecer el grado de dignidad con que fueron tratados los receptores.

Categoría	Excelente	Muy bueno	Bueno	Adecuado	Deficiente
Calificaciones de la dimensión	Más del 90 % de acuerdo con un grado alto o considerable	80-90 % de acuerdo con un grado alto o considerable	60-80 % de acuerdo con un grado alto o considerable, y no más del 15 % de acuerdo con un grado limitado o muy limitado	40-60 % de acuerdo con un grado alto o considerable, y no más del 15 % de acuerdo con un grado limitado o muy limitado	Menos del 40 % de acuerdo con un grado alto o considerable

Fuente: Basado en Oakden (2013: 10)

Las calificaciones de diferentes dimensiones pueden ser consolidadas en una calificación general del elemento, tal como se muestra a continuación.

Calificación general del elemento	Excelente	Muy bueno	Bueno	Adecuado	Deficiente
<b>Calificación de las dimensiones</b>	Todas las dimensiones calificadas como excelentes	Casi todas las dimensiones calificadas como muy buenas o excelentes; ninguna dimensión calificada como deficiente o adecuada	La mayoría de las dimensiones calificadas como buenas o mejores; ninguna calificada como deficiente; más dimensiones calificadas como muy buenas o excelentes que como adecuadas	La mayoría de las dimensiones calificadas como adecuadas o mejores; no más del 20 % calificadas como deficientes, y más dimensiones calificadas como mejores que adecuadas	Más del 20 % de las dimensiones calificadas de deficientes

El mismo enfoque puede utilizarse para consolidar las calificaciones de los diferentes elementos en un juicio de evaluación general de la actividad.



#### Consejo útil

Desarrollen, o inviten a su equipo a desarrollar, rúbricas de evaluación en un taller con las partes interesadas.

Llevar a cabo un taller con las partes interesadas con el objeto de desarrollar una rúbrica para medir la oportunidad de la actividad reduce las probabilidades de que surjan discrepancias en torno a la rúbrica, en comparación con la decisión de los evaluadores o el gestor de la evaluación de establecer dicha rúbrica arbitrariamente.

Davidson (2014) ofrece una excelente descripción general del razonamiento evaluativo. Si bien no se encontraron ejemplos del uso de rúbricas evaluativas en el sector humanitario al momento de redactar esta guía, Oakden (2013) menciona un caso en que fueron usadas. Las rúbricas evaluativas representan una poderosa herramienta para contrarrestar el eterno problema de falta de aprobación de los juicios de evaluación en la EAH porque se les considera meramente subjetivos.



#### Consejo útil

Utilicen rúbricas evaluativas, o inviten a su equipo a utilizarlas, cuando estimen posible que algunas partes interesadas impugnen las respuestas a las preguntas evaluativas argumentando que son subjetivas o se basan en anécdotas o sesgos.

## 16.4 Análisis de evidencia para preguntas causales

Las preguntas causales (ver la [Sección 6, Selección de las preguntas de evaluación](#)) indagan sobre las relaciones causales entre las intervenciones y los efectos. La causalidad constituye un problema en la EAH porque en la mayoría de las intervenciones intervienen muchos actores y los cambios son influenciados por múltiples factores, en vez de uno solo. Es por eso que la EAH normalmente se refiere a la contribución, antes que a la atribución.

La inferencia causal constituye la piedra angular de la [evaluación de impacto](#).



### **Definición: Inferencia causal**

Inferencia causal significa establecer una relación entre una causa y un efecto.

Es preciso entender que, como señalan Cook et al., “la inferencia causal, incluso en los experimentos, es fundamentalmente cualitativa” (2002: 6). Decidir lo que significa la relación entre determinadas variables ya es un juicio cualitativo, incluso si se sustenta en evidencia cuantitativa. Por ejemplo, decidir qué factores deben ser revisados porque posiblemente generan confusión es un juicio cualitativo.

Davidson (2009) ofrece ocho estrategias para establecer la inferencia causal:

1. Preguntar a quienes hayan visto o experimentado la causalidad de primera mano. En el diálogo Cook–Scriven sobre la causalidad en la evaluación (Cook et al., 2010), Scriven deja muy claro que las personas pueden observar la causalidad de primera mano. Cook puntualiza que es necesario descartar otras explicaciones alternativas. En la EAH, si ustedes observasen que los comerciantes del mercado reducen los precios de los alimentos tan pronto como se anuncia una distribución de alimentos, tendrían buenas razones para inferir que el anuncio originó la reducción, después de haber hablado con los comerciantes para descartar otras explicaciones alternativas.
2. Verificar si el contenido de la intervención guarda relación con la naturaleza del efecto. Por ejemplo, si hubiese habido una reducción de la desnutrición severa después de la distribución de colchones, sería absurdo concluir que dicha distribución generó la reducción.
3. Buscar patrones distintivos en los efectos (el método del modus operandi) mientras buscan y eliminan explicaciones alternativas.
4. Verificar si la oportunidad de los efectos tiene sentido. Si B ocurre invariablemente después de A, en ese caso es probable que A lleve a B.
5. Examinar la relación entre la “dosis” y la “respuesta”. Por ejemplo,

- ¿demuestran menos síntomas los niños que han tenido la misma exposición a un acontecimiento traumático si han recibido un mayor número de sesiones de terapia que los que han recibido menos sesiones de terapia durante el mismo intervalo?
6. Utilizar un grupo de comparación o control (un diseño experimental o cuasiexperimental).
  7. Utilizar modelos estadísticos para eliminar otros factores causales posibles. Se trata de un enfoque utilizado en los estudios estadísticos cuasiexperimentales que emplean técnicas como estratificación, modelos multivariados, logística y regresión, así como análisis de la co-varianza, para excluir la influencia de variables que no son de interés. Ver Pourhosingholi et al. (2012) para una breve descripción de dichas técnicas.
  8. Identificar y revisar los mecanismos causales. Para ello pueden guiarse por la teoría del cambio (TdC) y verificar la evidencia para cada eslabón de la cadena de causalidades. Ver Funnell y Rogers (2011) para una descripción de las teorías de cambio y su utilización.

Ver Alexander y Bonino (2015) para una discusión de la causalidad en la EAH. Rogers (2014) brinda una buena descripción general de la causalidad. Befani (2012) ofrece una revisión académica exhaustiva como parte del estudio de Stern et al. (2012).

## 16.5 Análisis de evidencia para responder las preguntas descriptivas

Las preguntas descriptivas (ver la Sección 6, Selección de las preguntas de evaluación) en ocasiones requieren el uso de métodos estadísticos (ver la Sección 16.6) para describir la intervención identificando promedios, etc. El análisis de la información cualitativa codificada también puede ser utilizado para identificar los temas y problemas centrales de la intervención.

### Análisis cualitativo de información primaria

El análisis de información cualitativa debe ser un proceso riguroso y lógico para convertir la información de la evaluación en hallazgos. El análisis formal de la información cualitativa normalmente depende de la adjudicación de categorías a determinadas piezas de evidencia, y luego de la agrupación de dichas piezas. La asignación de estas categorías se denomina codificación.

**Definición: Codificación**

Codificación significa asignar categorías a piezas de evidencia específicas.

Las categorías se describen por medio de códigos, que pueden consistir de encabezados simples describiendo la categoría o, en esquemas de codificación más complejos, de un conjunto de caracteres alfanuméricos. En la EAH, la codificación casi siempre se limita al uso de encabezados para recordar la categoría al codificador.

**La codificación en la EAH**

En su expresión más simple, codificar puede consistir en leer los apuntes de campo y etiquetar diferentes categorías de evidencia, o simplemente en registrar los apuntes de campo de acuerdo a categorías establecidas. Entre los enfoques simples se encuentran los siguientes:

- Leer los apuntes y adherir etiquetas (códigos) a la evidencia de la categoría relacionada;
- Cortar y pegar los apuntes en una nueva estructura organizada por temas;
- Marcar los documentos físicos con marcadores o pegatinas de colores;
- Añadir etiquetas de texto, comentarios o subrayados a los documentos electrónicos.

**En profundidad: La codificación en la EAH**

Las evaluaciones más complejas requieren enfoques más rigurosos, como codificar los apuntes de campo. En la EAH, los códigos suelen ser palabras clave, y rara vez se utilizan códigos complejos. Una pieza de evidencia puede ser una única observación, una oración o frase de un informe o entrevista, una respuesta a una encuesta, un resumen de una encuesta o cualquier otra afirmación de hecho u opinión. Típicamente, una entrevista semiestructurada de 45 minutos de duración puede generar entre 10 y 20 piezas de evidencia. Saldaña (2012: 1-41) ha escrito una excelente introducción al tema.

En vista de las limitaciones presupuestarias y de tiempo, la codificación posterior al trabajo de campo es altamente inusual en la EAH. Una excepción es la evaluación de CARE de los esfuerzos de preparación y mitigación de desastres en Camboya, donde se involucró a todo el equipo en la “traducción y codificación de la información cualitativa y en la redacción del borrador del informe” (Ramage et al., 2006). El patrón más común en la EAH parece consistir en que los miembros del equipo de evaluación escriban secciones del informe con base en sus apuntes del trabajo de campo, o que el jefe del equipo de evaluación lo haga basándose en los informes escritos de los integrantes del equipo así como en sus propios apuntes.





Una práctica idónea actual en la EAH consiste en codificar la información cualitativa a medida que se va recogiendo. Los códigos iniciales son determinados por las preguntas de evaluación y por los temas que emergen durante la fase inicial. Una evaluación humanitaria típica puede tener entre 20 y 30 códigos. Se pueden extraer códigos adicionales de (según Miles et al., 2013):

- El marco conceptual (ver la Sección 5, Enmarcando su evaluación). Por ejemplo, si estuviésemos usando el enfoque de medios de vida sostenibles como nuestro marco conceptual, tendríamos códigos para: vulnerabilidad, los distintos activos, estructuras, procesos, estrategias de generación de ingresos y efecto en los medios de vida.
- Las áreas problemáticas. Las áreas problemáticas comunes en la acción humanitaria son: focalización, género y coordinación.
- La hipótesis. Si estamos manejando la hipótesis de que la ayuda monetaria ha generado conductas antisociales, en ese caso tendríamos códigos para los distintos tipos de conductas antisociales, y los códigos incluirían evidencia de la ausencia de dichas conductas.
- Los problemas centrales. En un proyecto de alojamiento, los problemas centrales podrían incluir los niveles de renta, la ocupación de un albergue por más de una familia, etc.

El cuadro de códigos que aparece a continuación revela los códigos de la ETR interinstitucional de la respuesta a la crisis en el Valle de Swat, en Pakistán (Cosgrave, 2009). Los códigos fueron derivados de: una pregunta de evaluación, áreas problemáticas comunes, problemas centrales, problemas mencionados en los TdR o problemas que emergieron durante el trabajo de campo (tales como: neutralidad, acceso, burocracia, retrasos y factores contextuales). Varios de los códigos fueron derivados de múltiples fuentes, pero se menciona únicamente la fuente principal.

Código	Fuente del código	Número de piezas de evidencia	Número de fuentes distintas
<b>Cambios de 2009 a 2010</b>	Problema mencionado en los TdR	57	23
<b>Conectividad</b>	Pregunta de evaluación	64	37
<b>Reconstrucción temprana</b>	Problema mencionado en los TdR	41	26
<b>Enfoque de género</b>	Área problemática	26	16



Código	Fuente del código	Número de piezas de evidencia	Número de fuentes distintas
<b>Gobierno</b>	Problema mencionado en los TdR	53	24
<b>Aprendizaje para el futuro</b>	Problema mencionado en los TdR	61	24
<b>Militares</b>	Pregunta de evaluación	35	16
<b>Seguimiento</b>	Problema mencionado en los TdR	23	16
<b>Neutralidad</b>	Trabajo de campo	92	36
<b>Revisiones</b>	Pregunta de evaluación	7	5
<b>Logros</b>	Pregunta de evaluación	81	39
<b>Financiamiento</b>	Problema central	91	36
<b>Coordinación</b>	Problema central	144	39
<b>Acceso</b>	Trabajo de campo	32	21
<b>Cambios en el aprendizaje</b>	Problema mencionado en los TdR	22	12
<b>Seguridad</b>	Área problemática	56	30
<b>Evaluaciones</b>	Pregunta de evaluación	43	32
<b>Campamentos-Anfitrión-Otros</b>	Problema mencionado en los TdR	58	34
<b>Consulta</b>	Pregunta de evaluación	12	11
<b>Brechas</b>	Área problemática	109	47
<b>Focalización</b>	Área problemática	82	45
<b>Burocracia</b>	Trabajo de campo	38	21
<b>Retrasos</b>	Trabajo de campo	18	14
<b>Problemas contextuales</b>	Trabajo de campo	57	29
<b>Total</b>		<b>1.306</b>	<b>160 en total</b>





La codificación se lleva a cabo revisando la información recogida y abstrayendo la evidencia:

1. Entrevistas y otros apuntes: El equipo de evaluación ingresa los apuntes de sus entrevistas en la herramienta de evidencia, codificándolos en ese mismo instante.
2. Documentos físicos: Se ingresan segmentos clave en la herramienta de evidencia.
3. Documentos electrónicos: Se copian, depuran y pegan segmentos clave en la herramienta de evidencia (la depuración es necesaria para eliminar los retornos automáticos de carro y los códigos de formato del texto.)
4. Informes de encuestas: Para encuestas en línea el proceso puede ser automatizado si las preguntas de la encuesta están alineadas con los códigos. El resultado se revisa manualmente y se añaden códigos adicionales según proceda.

La codificación se hace mucho más fácil si todas las herramientas de recolección de información, guías de entrevistas, formularios de encuestas, etc., están alineados con los códigos iniciales.



#### Consejo útil

Utilicen las categorías iniciales para organizar toda la recolección de información ya que, si las categorías se modifican más adelante, es más fácil hacer los ajustes correspondientes si la información está organizada que si está desorganizada.

Utilicen una herramienta de evidencia simple basada en una hoja de cálculo para facilitar el proceso de codificación, que puede ser cotejado automáticamente para obtener un solo resumen de la evidencia.

Código	Evidencia	Fuente	Iniciales	2° código	3° código
<b>El código del tema para esta pieza de evidencia</b>	Detalles sobre la pieza de evidencia (normalmente un retazo de texto o <i>snippet</i> ), que pueden abarcar desde una oración hasta un párrafo. El conteo de palabras promedio para una pieza de evidencia en el ejemplo que aparece arriba era de 22.	Un código que alude a la fuente específica, trátase de una entrevista, reunión, documento u observación	Quién escribió esto	Un segundo código para el mismo texto	Un tercer código para el mismo texto



Aunque la herramienta puede usarse con un código único para cada pieza de evidencia, se pueden añadir códigos adicionales para la misma pieza de evidencia de ser necesario. Si hay más de tres códigos para una misma pieza de evidencia, el texto puede ser reingresado con los códigos adicionales.

Las fuentes de evidencia pueden ser variadas. En el caso de una ETR realizada en Filipinas, más del 40 por ciento de la evidencia provino de comentarios específicos a las preguntas abiertas de una encuesta en línea (Cosgrave y Ticao, 2014). Esta, combinada con entrevistas telefónicas, tuvo un papel destacado en la evaluación, ya que muchos de los informantes clave más importantes ya habían abandonado Filipinas.

Metodología de recolección de información	Piezas de evidencia	Como %
Encuesta en línea	944	42.1 %
Entrevistas telefónicas	523	23.4 %
Entrevista semiestructurada (dos o más entrevistados)	390	17.4 %
Entrevista semiestructurada (un solo entrevistado)	230	10.3 %
Reunión general	117	5.2 %
Comentarios enviados por correo electrónico	21	0.9 %
Discusión detallada (>10 minutos sobre uno o más temas)	11	0.5 %
Observación	2	0.1 %
<b>Total de piezas únicas de evidencia</b>	<b>2.238</b>	<b>100 %</b>

### Extrayendo conclusiones sin el software CAQDAS

El equipo de evaluación utiliza los códigos para ordenar toda la información y cotejarla por códigos. El jefe del equipo, o un miembro del equipo asignado a códigos específicos, revisa luego toda la información recolectada bajo un mismo tema, antes de extraer los hallazgos. Al considerar toda la información codificada, inevitablemente surgirán algunos conflictos.

Los conflictos en torno a la evidencia pueden surgir respecto de la información cuantitativa tanto como respecto de la cualitativa. Un ejemplo de conflicto relacionado con evidencia cuantitativa es la falta de coincidencia entre las

cantidades mensuales distribuidas y las cantidades mensuales despachadas desde los almacenes – las razones para ello pueden ser complejas e incluyen: despacho y distribución en meses distintos, definiciones diferentes, almacenamiento local en los lugares de distribución; gestión de discrepancias, devoluciones o pérdidas en los lugares de distribución, etc.

Al considerar las controversias en torno a la evidencia:

- Atribuyan mayor peso a las opiniones de los más afectados. Por ejemplo, las opiniones de los beneficiarios en cuanto a si fueron consultados o no tienen un peso mucho mayor que las opiniones del personal de la organización en relación con el mismo tema. Las opiniones de los beneficiarios que han sido pasados por alto debido a deficiencias de coordinación deben recibir un mayor peso que las opiniones de las personas que participaron en las reuniones de coordinación.
- Las opiniones de una persona con considerable experiencia en el sector podrían tener un mayor peso que las de una persona con poca experiencia. Sin embargo, ese no siempre es el caso. Una persona con gran experiencia puede obsecarse respecto de un enfoque o una actitud en especial.
- Consideren adjudicar un mayor peso a los puntos de vista que parezcan contradecir el interés propio del entrevistado. Por ejemplo, presten mucha atención a los beneficiarios cuando afirman que ya no desean recibir un tipo de asistencia determinado. Del mismo modo, presten estrecha atención a las críticas del personal de la organización a su propio programa, especialmente si no proyectan otras señales de descontento.
- Otorguen mayor peso a aquellos puntos de vista que triangulen otra evidencia.

Este enfoque garantiza que los hallazgos del informe de evaluación se sustenten fuertemente en evidencia recolectada. En caso surjan controversias significativas en torno a la evidencia, estas deberán ser reportadas, junto con el razonamiento utilizado para afrontarlas.

### El software CAQDAS para codificar la información después del trabajo de campo

La codificación posterior al trabajo de campo puede ser un proceso lento y rara vez se utiliza en la EAH. Sin embargo, existen herramientas informáticas poderosas que pueden ser de utilidad, por ejemplo los programas de Análisis Cualitativo de Datos Asistido por Computadora o CAQDAS (por sus siglas en inglés). Sin embargo:

- Se trata de un software costoso;
- La codificación exige una gran cantidad de tiempo;
- El funcionamiento de los paquetes de software es difícil de captar con rapidez.

Por esta razón, los paquetes de CAQDAS rara vez son utilizados en la EAH. LaPelle (2004) demuestra que los procesadores de texto genéricos de Word pueden usarse para ejecutar las tareas básicas del CAQDAS. Koenig (2004) ofrece un muy buen resumen práctico del software CAQDAS, incluyendo potenciales escollos.

Miles et al. (2013) advierten que “algunos programas CAQDAS contienen características maravillosas para respaldar los análisis, pero suponen una curva de aprendizaje empinada en la mayoría de los casos”. Silver y Lewins (2014) han desarrollado una guía pormenorizada para el manejo de algunos de los paquetes más populares. Independientemente de lo anterior, Baugh et al. (2010) describen el tiempo requerido para aprender a usar un paquete CAQDAS como una de las mayores limitaciones a la hora de utilizar dichos programas.

### **Análisis cualitativo de información secundaria**

Este tema ya ha sido parcialmente tratado en la [Sección 10, Métodos documentales](#). El objetivo principal aquí es hacer hincapié en que la información secundaria también puede ser ingresada en la herramienta de evidencia. El trabajo de campo principal de la evaluación puede dar acceso a información secundaria adicional que necesitará ser analizada utilizando los métodos descritos para los estudios documentales.

#### **Análisis de portafolio**

El análisis de portafolio es una herramienta poderosa para examinar la estrategia revisando la información del proyecto. En su expresión más simple, el análisis de portafolio puede consistir de apenas un análisis descriptivo de aspectos tales como qué porcentaje de la ayuda fue entregado a qué contrapartes o sectores, etc. Además, puede examinar los tipos de subvenciones promedio por cantidades. Ese es el tipo de análisis de portafolio que fue realizado por la evaluación de la estrategia humanitaria danesa (Mowjee et al., 2015) y por la evaluación conjunta del apoyo a los esfuerzos de construcción de la paz en Sudán del Sur (Bennett et al., 2010).

A un nivel ligeramente más profundo, puede examinar qué tipos de proyectos, sectores o contrapartes han sido menos propensos a solicitar una extensión del presupuesto o hallar problemas.

El análisis de portafolio puede implicar asimismo el uso de una herramienta de calificación o un sistema de indicadores resumido, como en el caso de la evaluación australiana de la asistencia de las ONG tras el terremoto de Pakistán en octubre de 2005 (Crawford et al., 2006). Puede implicar también la calificación de proyectos individuales.

Un análisis de portafolio más profundo puede consistir en desarrollar una serie de herramientas de calificación y luego calificar cada proyecto de la cartera de proyectos con base en ellas. Este fue el enfoque adoptado por la evaluación quinquenal del CERF (Channel Research, 2011) de las áreas de género, vulnerabilidad y otros temas transversales.

## 16.6 Análisis estadístico de información primaria

Esta sección no pretende ser una presentación exhaustiva del análisis estadístico. Se trata de un tema altamente técnico y existe una gran variedad de publicaciones sobre el tema que duplican con creces el tamaño de esta guía.

Cabe anotar que la información recolectada a través de encuestas no solo es información numérica: también puede incluir información sobre categorías.

Las cuatro categorías de información son:

- Información sobre categorías
  - Nominal – por ejemplo, el género (mujer/hombre)
  - Ordinal – en que las categorías son jerarquizadas (nivel socioeconómico, por ejemplo, o “muy insatisfecho, insatisfecho, satisfecho y muy satisfecho”)
- Información numérica
  - Datos a intervalos – información numérica para la cual no hay un cero real (por ejemplo, para calificar la seguridad alimentaria)
  - Datos de ratios – información numérica en la cual hay un cero significativo y donde los números pueden ser multiplicados de manera significativa (por ejemplo, el peso ganado en un programa de alimentación)

Los procedimientos estadísticos utilizables varían según el tipo de información.

### Depuración e ingreso de la información

La recolección electrónica de información simplifica enormemente la depuración y el ingreso de la información. Sin embargo, la información aún tiene que ser revisada, ya que los programas de computación no corrigen todos los errores.

La recolección electrónica de información ha puesto fin en gran medida a la que solía ser la onerosa tarea de ingresar la información.

### Información faltante

El problema con la información faltante (de encuestados que no responden o responden solo parcialmente) es que siempre existe la preocupación de que aquellos que no respondieron sean de alguna manera distintos a los que sí lo hicieron, y que esto pueda sesgar los resultados. Se trata de una preocupación válida, en especial en las encuestas por internet, pero la falta de respuesta también es un problema en las encuestas cara a cara. Las razones de la falta de respuesta pueden incluir:

- La persona a entrevistar no estaba disponible (la encuesta puede tener un protocolo para nuevos intentos o encuestados alternativos);
- Negativa a responder ciertas preguntas;
- Incapacidad para responder ciertas preguntas (no sabe);
- Ser pasado por alto por el encuestador.

Un buen diseño y validación del cuestionario pueden ayudar a reducir estas últimas tres categorías. La primera es una fuente de preocupación especial porque un potencial encuestado puede estar ausente por razones que produzcan un sesgo en los resultados de la encuesta. Por ejemplo, si un encuestador descubriera que el 10 por ciento de la muestra seleccionada no estaba disponible cuando los encuestadores los visitaron, la razón podría ser que estaban ocupados en alguna actividad importante fuera de la aldea. Excluir a los ausentes podría sesgar la encuesta, ya que se podría estar excluyendo a un grupo con un oficio o actividad específica y distintiva. Nuevamente, la validación previa sirve para identificar problemas como este.

## Estadística

El uso de la estadística en la EAH es de dos tipos – descriptivo e inferencial.



### **Definición: Estadística descriptiva**

La estadística descriptiva se utiliza para resumir los aspectos centrales de una población.

Las estadísticas descriptivas más comunes son las mediciones de la tendencia central (comúnmente denominadas promedios) y miden la dispersión o variabilidad (ver la [pág. 313](#)). Dichas estadísticas con frecuencia se utilizan para dar respuesta parcial a las preguntas de evaluación de índole descriptiva. Son útiles también para responder a las preguntas normativas, en especial cuando la norma o criterio en cuestión tiene un indicador numérico relacionado. Por ejemplo, la respuesta a la pregunta “¿Hasta qué punto cumplimos con las normas mínimas del Proyecto Esfera en lo que se refiere a la cantidad de agua suministrada?” probablemente debería incluir información acerca de la cantidad media suministrada por persona por día, así como el rango, antes de una discusión, sobre el grado al que dicho nivel de suministro cumplió con la norma de “una cantidad de agua suficiente para beber, cocinar y para higiene doméstica y personal”.



### **Definición: Estadística inferencial**

La estadística inferencial se utiliza ya sea para hacer deducciones acerca de una población a partir de una muestra o para hacer deducciones acerca de una hipótesis.

La estadística inferencial se utiliza no solo para responder preguntas descriptivas (tales como: ¿Qué porcentaje de la población afectada utilizó la mayor parte de su subsidio en efectivo para financiar los costos del alojamiento?) sino también para responder a preguntas causales (tales como: ¿En qué medida redujo el programa contra la desnutrición aguda, gestionado por la comunidad, los niveles de desnutrición severa y aguda?).



### Consejo útil

Indiquen siempre el nivel de confianza y el intervalo para las estadísticas inferenciales.

Por ejemplo, si el 50 % de la muestra invirtió la mayor parte de su subsidio en pagar el alojamiento, pero debido al tamaño reducido de la muestra el intervalo para un nivel de confianza del 95 % fue del 20 % a 80 %, se trata de un dato muy pertinente, ya que es 95 % seguro que entre un 20 % y un 80 % de la población haya utilizado la mayor parte de su subsidio en alojamiento.

### Tendencia central

La medida más común de la tendencia central para la información numérica probablemente sea la media (normalmente denominada el promedio), por ejemplo, el tamaño promedio de la familia o el tiempo promedio de permanencia en un campamento de tránsito. La tendencia central a menudo se utiliza para responder preguntas descriptivas, y puede describir el número o volumen promedio de los subsidios en efectivo distribuidos, etc. Utilizar este tipo de promedios puede dar una idea de la magnitud de las intervenciones rápidamente.

La media es una medición de la tendencia central. Otras dos mediciones posibles de la tendencia central son: la mediana, que es el valor central de un conjunto de datos en el que la mitad de los valores es grande y la mitad es pequeño, y la moda estadística, que es el valor más común.

El valor mediano con frecuencia puede ser una medición más útil de la tendencia central.

### Nuevos refugiados en un campamento para refugiados ya existente

Consideren un campamento para refugiados que tiene una población de 2,000 personas que llevan 20 años viviendo ahí. Una nueva afluencia de 8,000 personas buscó refugio en el campamento hace un mes. La extensión promedio del tiempo que los refugiados permanecen en el campamento es de  $(2,000 \times 20 + 8,000 \times 1) / 10,000 = 48.8$  meses – más de cuatro años. Sin embargo, la mediana estadística es de solo un mes (si organizaran a todos los refugiados por fecha de llegada, los números 4,999 y 5,000 hubieran permanecido ahí por un mes). La moda estadística (el tiempo más común de permanencia en el campamento) es también de un mes.

### Dispersión

Las mediciones de la dispersión (o variabilidad) demuestran cómo se distribuyen los valores alrededor de la media o mediana y pueden ser muy importantes para responder preguntas de corte descriptivo. Por ejemplo, afirmar que el subsidio promedio para actividades de apoyo a medios de vida era de \$300 sería una descripción deficiente si no se menciona al mismo tiempo que el rango oscilaba

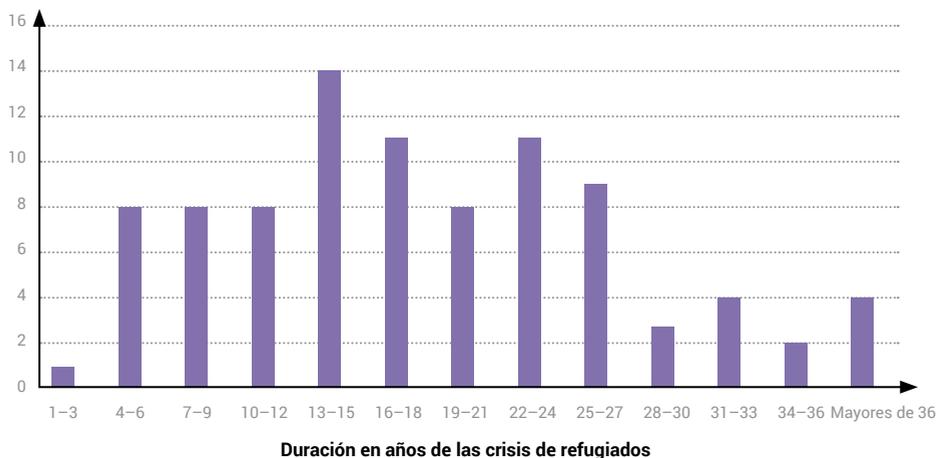
entre \$30 y \$600, pero que el 90 % de los receptores recibieron entre \$200 y \$400. Las opciones incluyen mencionar:

- El rango – la diferencia entre los valores más altos y más bajos (por ejemplo, un tamaño de familia de un rango que oscilaba entre uno y 19 miembros);
- Los subrangos – la diferencia entre los valores más altos y más bajos aplicables a un grupo en particular (los hogares monoparentales recibieron un subsidio de \$150 a \$450);
- El rango intercuartílico – la diferencia entre los valores que representan el 25 % y el 75 % de los valores de la distribución;
- La varianza – la suma de los cuadrados de la diferencia entre cada valor y la media;
- La desviación típica o desviación estándar – la raíz cuadrada de la varianza. Este valor tiene la ventaja de estar en la misma unidad que la media, de modo tal que, si la medición es en años, la desviación típica estaría en años también.

Sin embargo, la varianza y la desviación típica son mediciones bastante técnicas. Las tablas de distribución con frecuencia son más informativas para los usuarios primarios previstos. El gráfico que mostramos a continuación presenta la información de manera más sencilla que una afirmación sobre la duración media del desplazamiento y la desviación típica.

**Figura 16.2:** Distribución de la frecuencia de las crisis de refugiados en 91 países de origen en el período 1978–2014, con más de 5,000 refugiados en el momento más álgido (excluyendo las crisis recientes de Siria y Sudán del Sur)

#### Número de países



Fuente: Elaboración propia de la autora

## Validación de la hipótesis estadística

La principal ventaja de recolectar información *big-n* es que permite validar las hipótesis por medio de la estadística. Las hipótesis incluyen:

- Hipótesis sobre las características de un grupo específico con base en las características de una muestra seleccionada aleatoriamente (por ejemplo, estimando la media poblacional);
- Hipótesis sobre la diferencia o relación entre dos grupos (por ejemplo, entre el grupo asistido y un grupo de comparación que no recibió asistencia, o entre el grupo asistido antes y después de la intervención).

El enfoque habitual consiste en validar la diferencia entre la hipótesis y la hipótesis nula. Si, por ejemplo, la hipótesis es que la población que recibió un kit con productos de albergue hace un año tenía una mayor seguridad alimentaria en el hogar que un grupo equivalente que no recibió el kit, la hipótesis nula sería que no tiene una puntuación más alta en seguridad alimentaria en el hogar.

La hipótesis nula es similar al contrafactual, que es la situación que hubiera prevalecido si no se hubiera llevado a cabo la asistencia. En el caso que nos ocupa, el grupo equivalente que no recibió el kit con artículos de albergue representa el contrafactual. En realidad, nunca es posible examinar de verdad el contrafactual (porque es imposible validar simultáneamente los resultados de proporcionar y negar asistencia a la misma persona al mismo tiempo), de modo tal que la hipótesis nula es una aproximación.

En la validación de la hipótesis podemos concluir que la hipótesis es verdadera o que es falsa. Si concluimos que es verdadera, podríamos acertar o estar equivocados. Si nos equivocamos en concluir que la hipótesis es falsa, se trata de un error falso positivo y convencionalmente utilizamos un valor del 5 % para eso: es decir, nuestra conclusión de que la hipótesis es verdadera tiene un 5 % de probabilidades de estar equivocada.

Si concluimos equivocadamente que la hipótesis es falsa, en ese caso se trata de un error falso negativo. Convencionalmente, se asume que este es del 20 %, pero ver la [Sección 12, Muestreo](#), donde se sugiere que una probabilidad del 10 % de un falso negativo es más apropiada.

### Pruebas estadísticas

Las pruebas utilizadas dependen del tipo de información (sobre categorías o numérica, tal como se describió anteriormente), así como del número de variables dependientes e independientes.

- Variable dependiente: el efecto de interés – en el ejemplo que nos ocupa, las puntuaciones para seguridad alimentaria en el hogar. Se trata de una variable numérica.
- Variable independiente: el factor de entrada de interés – en este caso, el suministro de kits con artículos de albergue. Se trata de una variable nominal (recibió kit/no recibió kit).

La selección de la prueba adecuada puede ser una tarea técnica compleja. Las pruebas estadísticas en sí dependerán también del diseño de evaluación. Se requieren análisis diferentes para diseños diferentes, es decir, diseños de regresiones múltiples, series temporales interrumpidas, regresión discontinua, etc. Los equipos de evaluación pueden acometer dichos diseños si poseen la capacidad técnica para analizarlos únicamente.

Como regla, deberíamos analizar todos los efectos aparentes para determinar su significación estadística. En este caso, por ejemplo, si encontrásemos que el grupo asistido tiene una puntuación de seguridad alimentaria en el hogar un 5 % más alta que la del grupo no asistido, deberíamos analizar la probabilidad de que este resultado haya ocurrido por azar.

Consulten un texto confiable para detalles sobre cómo llevar a cabo una prueba estadística. Muchos textos universitarios actuales sobre estadística parecen apoyarse en el uso de un paquete de software como SPSS.<sup>17</sup> Dichos paquetes son bastante caros. Microsoft Excel tiene capacidades estadísticas básicas, las mismas que pueden ser mejoradas con un complemento gratuito como Real Statistics (Zaointz, 2015), que añade una gama de funciones, entre ellas algunas que sirven para procesar información *big-n*.

## 16.7 Análisis numérico de información secundaria

El análisis numérico de información secundaria es similar al análisis de información primaria, pero más enfocado en estadísticas, tendencias y correlaciones descriptivas. La revisión sobre desplazamiento prolongado (Crawford et al., 2015) se sustentó en parte en el análisis de información secundaria sobre refugiados recolectada por ACNUR, pero utilizó también información secundaria sobre desplazamiento interno de las siguientes fuentes: Centro de Seguimiento de Desplazamiento Interno, información sobre desarrollo del PNUD, información económica del Banco Mundial, información sobre estabilidad del Índice de Fragilidad de Estados e información financiera de ACNUR, el PMA y UNICEF.

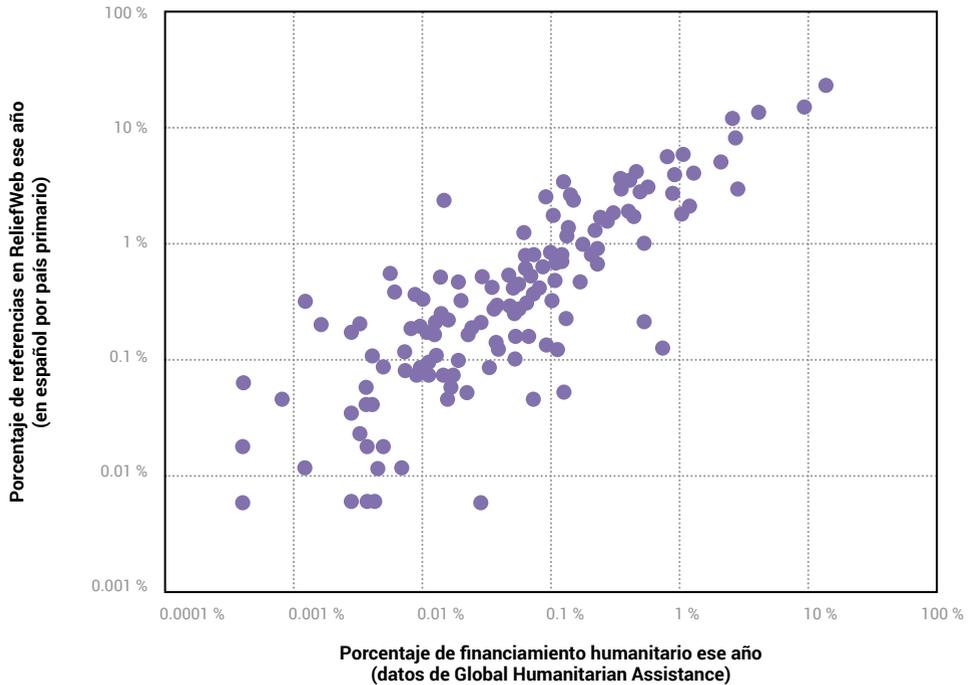


### Consejo útil

Generen siempre un diagrama de dispersión o *scatterplot* de cualquier información numérica que estén examinando.

Los diagramas de dispersión sacan a relucir las anomalías, revelan hasta qué punto los datos se correlacionan y pueden conducir a valiosas líneas de indagación. Cualquier programa de hojas de cálculo puede producir un diagrama de dispersión. Es importante recordar que la correlación no implica causalidad. Puede ser que la correlación se produzca íntegramente por azar, o que las variables aparentemente dependientes e independientes en realidad estén siendo influenciadas por una tercera o cuarta variable, etc.

**Figura 16.3:** Correlación entre el financiamiento humanitario global y las referencias en ReliefWeb en 2010 (cada punto representa un país)



Fuente: Elaboración propia de la autora

La revisión sobre desplazamiento prolongado (Crawford et al., 2015) utilizó diagramas de dispersión para analizar las correlaciones entre datos antes de la validación estadística. En este caso, el diagrama de dispersión estaba analizando la relación entre el financiamiento humanitario y las referencias a un país aparecidas en el portal web ReliefWeb. El interés estribaba en si la distribución de las referencias aparecidas en ReliefWeb podría ser utilizada como sustituto de la distribución del financiamiento humanitario durante los últimos años sobre los cuales aún no había información financiera disponible.

### Pruebas estadísticas comunes utilizadas en el análisis numérico

Cualquier equipo de evaluación debería ser capaz de realizar pruebas estadísticas para:

- Determinar la correlación. ¿Cuáles son las probabilidades de que una correlación se haya producido por azar?
- Pruebas de independencia, tales como las pruebas de chi-cuadrado. ¿Están relacionados dos factores de alguna manera o son independientes?

El software de hojas de cálculo común incluye las funciones necesarias para dichas pruebas. Estas pruebas estadísticas básicas son fundamentales para establecer si es probable o no que una correlación haya surgido por azar. Las pruebas de independencia son necesarias para establecer si dos factores son independientes antes de perder el tiempo investigando su relación.

Por ejemplo, en la evaluación del Consorcio de Agencias Humanitarias Británicas (Cosgrave y Polastro, 2012), el equipo utilizó una prueba de chi-cuadrado para constatar la relación entre la pertenencia al comité de financiamiento y el porcentaje de solicitudes de financiamiento que obtuvieron fondos. El impacto de la pertenencia al comité en el éxito de una solicitud no era estadísticamente significativo a un nivel del 5 %.

## 16.8 De la evidencia a las recomendaciones

Los informes de evaluación normalmente incluyen: evidencia, hallazgos, conclusiones y recomendaciones. La evidencia generalmente se ofrece para ilustrar cómo el equipo llegó a la conclusión acerca de un hallazgo en particular. La sumatoria de la evidencia recolectada utilizada en una evaluación incluye: los apuntes del equipo, toda la información recopilada y el conjunto de referencias recogido. Solo una fracción de lo anterior puede ser presentada en el informe.



### Consejo útil

Hagan un esfuerzo para que la evidencia presentada en el informe cuente. Incluyan citas de los beneficiarios y utilicen fotografías, cuadros y gráficos para ilustrar y resumir los aspectos centrales (cerciorándose de obtener el consentimiento y respetar la confidencialidad cuando así se requiera; ver la [Sección 14](#) para mayores detalles).



### Definición: Hallazgo

Un hallazgo es una afirmación factual basada en evidencia.



### Definición: Conclusión

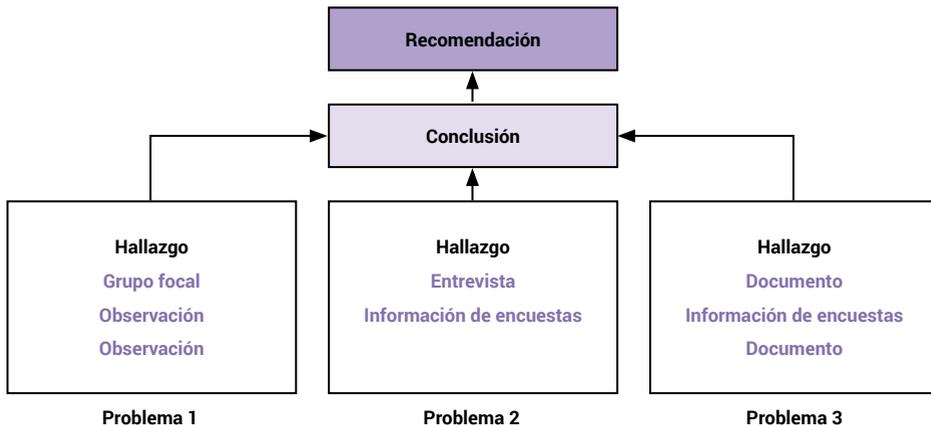
Una conclusión es una afirmación inductiva basada en uno o más hallazgos.



### Definición: Recomendación

Una recomendación es un curso de acción que los evaluadores sugieren como forma de responder a una o más conclusiones.

Es necesario que exista un flujo lógico o proceso de análisis claro desde los hallazgos, pasando por las conclusiones, hasta las recomendaciones. Esto debe ser presentado de la manera más transparente posible en los productos de evaluación. Por ejemplo, los hallazgos deben estar claramente respaldados por evidencia; las conclusiones, por hallazgos, y las recomendaciones, por conclusiones, en lo que asemeja a grandes rasgos la estructura de un informe científico.



Jane Davidson (2007) propone una forma diferente de organizar un informe de evaluación. La autora sugiere que las respuestas a las preguntas de evaluación deberían ser presentadas primero, seguidas por la evidencia y las conclusiones. Además, sugiere estructurar el informe en torno a las preguntas de evaluación e incluir todo lo relacionado con cada pregunta en la sección pertinente, incluidas las recomendaciones.

Las conclusiones pueden presentarse ya sea al final de cada sección o todas juntas en una sección final. Presentar las conclusiones al final de cada sección hace que permanezcan más cerca de los hallazgos relacionados y ayuda a transparentar el flujo lógico.

También es posible presentar una sección final con un resumen de las conclusiones.

## Recomendaciones

Las recomendaciones pueden ser:

- Omitidas del informe. La lógica detrás de limitarse a extraer conclusiones es que la organización que encarga la evaluación está mejor posicionada para traducir las conclusiones en recomendaciones. Este enfoque es poco habitual en la EAH, dado que pocas organizaciones, si acaso alguna, cuentan con la estructura necesaria para traducir las conclusiones en recomendaciones. La ETR de la respuesta de Plan Internacional al tifón Haiyan en Filipinas no incluyó recomendaciones en el informe final (Cosgrave y Ticao, 2014).
- Generadas por el equipo de evaluación y presentadas en el informe de evaluación. Este es el enfoque más común en la EAH, y la mayoría de los informes de evaluación de EAH contienen recomendaciones.
- Desarrolladas como resultado de un proceso de consulta con las partes interesadas. Nuevamente, este enfoque es poco común en la EAH. Los principios que subyacen a este enfoque son que alienta la apropiación de las recomendaciones y que las partes interesadas están óptimamente posicionadas para sugerir soluciones a los problemas identificados en las conclusiones. La evaluación conjunta de la respuesta al terremoto de Yogyakarta (Wilson et al., 2007) utilizó este enfoque (ver [ejemplo de buena práctica en la pág. 346](#)).
- Perfeccionadas en un proceso de consulta con las partes interesadas, a partir de recomendaciones preliminares inicialmente desarrolladas por el equipo de evaluación. Este fue el enfoque adoptado en la ETR de la respuesta a las inundaciones de Pakistán en 2010 (Polastro et al., 2011) (ver [ejemplo de buena práctica en la pág. 346](#)). La política de Acción contra el Hambre es que las recomendaciones deben ser genéricas, para que las partes interesadas las amplíen (ACF, 2011).

## Recomendaciones útiles

Las recomendaciones solo son útiles si:

- Son **específicas** – deben transmitir claramente qué es exactamente lo que se está recomendando.
- Están **enlazadas con acciones verificables** – deben permitir verificar si la recomendación ha sido implementada.
- **Identifican responsabilidades** – deben identificar a la persona o entidad responsable de poner en práctica la recomendación; la responsabilidad puede ser identificada con mayor precisión en la respuesta de la dirección al informe.
- Son **practicables** – las recomendaciones pueden implicar formas nuevas o poco habituales de hacer las cosas, pero deben tener siempre en cuenta la disponibilidad de recursos y otras restricciones.
- Tienen **plazos definidos** – deben incluir un cronograma para la ejecución de las recomendaciones, siempre que sea posible.
- Son **coherentes** – no deben contradecirse o parecer contradecirse entre sí.
- Figuran **en orden de prioridad** – debe desprenderse claramente qué recomendaciones son más importantes y cuáles son secundarias.
- Son **económicas** – las medidas recomendadas deben acarrear claramente beneficios proporcionales a los costos.

Asimismo, las recomendaciones deben ser limitadas en términos de cantidad. Cuanto menos recomendaciones contenga un informe, tanto más fácil será para la organización que encargó la evaluación ponerlas en práctica. Pero es posible que un evaluador experimentado haya detectado decenas de problemas de desempeño, en cuyo caso tiene la obligación ética de mencionarlos.

Existen diversas estrategias para ayudar a resolver este tipo de conflicto:

- Ofrecer una recomendación general y luego dar detalles de cómo podría ser implementada por diversos actores;
- Hacer recomendaciones menores de manera verbal e incluirlas en un anexo;
- Jerarquizar las recomendaciones en orden de importancia;
- Recomendaciones grupales agrupadas en torno al foco de la recomendación.

## Notas finales

### 11 / Diseños de evaluación para responder las preguntas de evaluación

1. El capítulo en cuestión se encuentra disponible en línea en: [us.sagepub.com/sites/default/files/upm-binaries/61527\\_Chapter\\_15.pdf](http://us.sagepub.com/sites/default/files/upm-binaries/61527_Chapter_15.pdf).
2. El desconocimiento no siempre es efectivo, pero los estudios rara vez validan este hecho (Hrobjartsson et al., 2007; Fergusson et al., 2004). Sackett (2007) apunta que, en tres estudios sobre la eficacia del desconocimiento, este fue exitoso solo en el 50 por ciento de los casos.

### 12 / Muestreo

3. La desviación estándar es la raíz cuadrada de la suma de los cuadrados de las diferencias entre los valores individuales de una variable y el valor medio de esa variable. La desviación estándar se encuentra expresada en las mismas unidades que la variable.

### 13 / Metodologías de campo

4. Tomado en parte del portal web de la Universidad Estatal de Michigan (Michigan State University) sobre preguntas de sondeo (Kennedy, 2006)
5. Se trata de una malla con el número de la entrevista en un eje y el número de adultos en el hogar en el otro. La intersección entre los dos da un número aleatorio para seleccionar al entrevistado (generalmente en orden de edad ascendente).
6. La revisión de NRC del programa de educación en Palestina (Shah, 2014: 72) utilizó un instrumento de observación estructurada para las actividades de observación en el aula. Este instrumento tenía 52 ítems de observación para medir cinco hipótesis subyacentes.
7. Ver: [www.mande.co.uk/docs/hierarch.htm](http://www.mande.co.uk/docs/hierarch.htm)
8. Ver: [betterevaluation.org/evaluation-options/richpictures](http://betterevaluation.org/evaluation-options/richpictures)



## Notas finales

### 14 / Involucrando a la población afectada en su evaluación

9. Ver la Red de Comunicación con Comunidades Afectadas por Desastres (CDAC, por sus siglas en inglés).
10. Casi el 75 % de las EAH evaluadas por ALNAP entre 2001 y 2004 fueron consideradas insatisfactorias o deficientes en “consulta e involucramiento de las partes primarias interesadas”, especialmente la población afectada (Beck y Buchanan-Smith, 2008).
11. Para orientación sobre cómo ejecutar este tipo de evaluación participativa, ver el trabajo de Better Evaluation en: [betterevaluation.org/plan/approach/participatory\\_evaluation](http://betterevaluation.org/plan/approach/participatory_evaluation), así como Alexander y Bonino (2014).
12. Ver Buchanan-Smith et al. (2015).
13. Esta subsección está muy basada en la Guía de Acompañamiento sobre Evaluación de la Protección (Bonino, 2016). El contenido ha sido adaptado aquí para que se relacione más ampliamente con la EAH. Para consideraciones específicas relacionadas con la evaluación de programas de protección, sírvanse consultar la Guía de Acompañamiento.

### 15 / Acceso restringido

14. Para una discusión en profundidad de lo que se ha llamado “contracción de espacio humanitario”, ver Collinson y Elhawary (2012).
15. Ver también Bush y Colleen (2015).

### 16 / Análisis

16. El Proyecto Esfera reemplazará las seis normas mínimas comunes a todos los sectores por las nueve Normas Humanitarias Esenciales (CHS Alliance, 2015) en las versiones futuras del Manual Esfera (Proyecto Esfera, 2015).
17. Un ejemplo es Andy Field (2006).

## Notas

11 / Diseños de evaluación para responder las preguntas de evaluación 192 - 213

12 / Muestreo 214 - 230

13 / Metodologías de campo 231 - 267

14 / Involucrando a la población afectada en su evaluación 268 - 282

15 / Acceso restringido 283 - 292

16 / Análisis 293 - 321

## Notas

# Comunicando y reportando los hallazgos y resultados

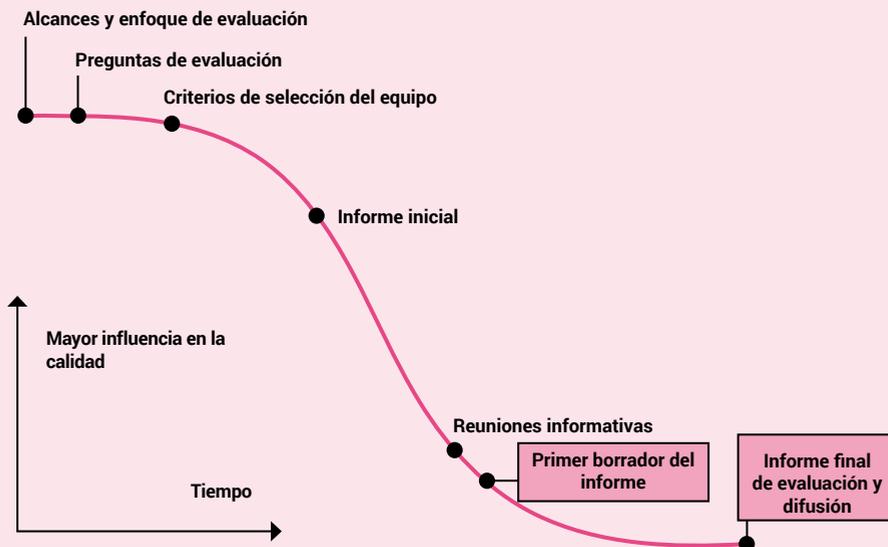


# 17 / Reportando y comunicando los hallazgos de la evaluación con enfoque de utilización

Como se señaló en la Sección 3, Pensar desde el principio y a menudo en el uso de la evaluación, al planear una evaluación es importante identificar lo que los usuarios primarios previstos necesitan saber y cómo comunicar esta información de la manera más idónea posible para facilitar su uso. ¿Cómo deberían ser presentados los hallazgos de la evaluación a los usuarios? Adicionalmente al informe de evaluación, ¿qué otros productos debemos hacerles llegar?

Esta sección ofrece orientación al equipo de evaluación para que plasmen sus hallazgos en un informe de alta calidad, atractivo y accesible. A ello le siguen algunas ideas para presentar los hallazgos de la evaluación con la intención de maximizar su uso. Por último, se ofrece orientación para que el gestor de la evaluación y la organización garanticen la calidad del informe de evaluación durante el proceso de aprobación del mismo, así como en las actividades de difusión y comunicación una vez que el informe esté listo.

Puesto de control de calidad: Primer borrador del informe, informe final de evaluación y difusión



Esta sección incide en dos puntos centrales a nivel de control de calidad. En primer lugar, el informe en borrador representa una oportunidad para que el gestor de la evaluación verifique la calidad del análisis y ejerza influencia en la forma de presentar los resultados en el informe final. En segundo lugar, ya con el informe de evaluación en mano, el gestor de la evaluación puede dedicarse ahora a difundir los hallazgos y resultados, adecuando la presentación para que se ajuste a las necesidades de las diferentes partes interesadas.

## 17.1 Productos centrales

Los productos requeridos del equipo de evaluación normalmente incluyen: informe inicial, talleres informativos, asesoría provista directamente en terreno e informe de evaluación. El cuadro que presentamos a continuación ofrece un resumen de los principales productos de evaluación, la oportunidad de los mismos y los procesos de aprobación. Es posible que el equipo de evaluación deba participar en las actividades de difusión.

Cuadro 17.1: Principales productos de evaluación

Producto	Oportunidad	Proceso de revisión y aprobación
<u>Informe inicial</u>	Al menos dos semanas antes del trabajo de campo, para dejar un espacio para los comentarios y ajustes a las metodologías y planes de trabajo de campo	Circulación del documento entre una selección de partes interesadas (generalmente las partes interesadas que tienen un interés directo en la evaluación y/o los usuarios primarios previstos), solicitando sus comentarios  Reunión entre el grupo referencial o el grupo de coordinación y el equipo de evaluación
Asesoría provista directamente en terreno	A medida que surgen problemas sobre el terreno y en respuesta a las observaciones, especialmente en las ETR	Ninguno
Talleres para informar sobre el trabajo de campo	Hacia la mitad de una ETR, para que los miembros del equipo de evaluación puedan validar los hallazgos emergentes  Al término del trabajo de campo en otros tipos de EAH, generalmente para presentar los hallazgos preliminares y validarlos	Comentarios de las partes interesadas a los hallazgos y conclusiones preliminares de los evaluadores, ya sea validándolos o cuestionándolos
Nota informativa sobre el trabajo de campo	Al término del trabajo de campo (puede ser tan sencilla como la presentación del taller para informar sobre el trabajo de campo)	La versión final debe reflejar la retroalimentación recibida en la reunión informativa sobre el trabajo de campo así como los hallazgos preliminares.
Borrador del informe de evaluación	Tres a cuatro semanas después de finalizar el trabajo de campo, más en evaluaciones complejas; puede pasar por varias revisiones	Revisión inicial de la calidad del borrador por el gestor de la evaluación  Circulación del borrador al grupo referencial o al grupo de coordinación, incluidas las partes interesadas que tienen un interés directo en la evaluación y los usuarios previstos, solicitando sus comentarios  Se puede celebrar una reunión o un taller entre el grupo referencial o el grupo de coordinación y el equipo de evaluación, para conversar sobre aspectos centrales.
Informe final de evaluación	Aproximadamente dos semanas después de recibir los comentarios finales al penúltimo borrador	Debe ser aprobado por el gestor de la evaluación y/o el grupo de gestión más amplio.

## 17.2 Informes de evaluación

Se trata del principal instrumento para presentar los hallazgos, conclusiones y recomendaciones de la evaluación. Como dice la nota “cómo hacerlo” de USAID sobre la elaboración de los informes de evaluación de USAID (USAID, 2012: 1): Un factor clave en el uso de los hallazgos de una evaluación es contar con un informe bien escrito, conciso, que comunique con claridad y rapidez hallazgos y conclusiones creíbles e incluya gráficos fáciles de entender y un formato coherente.

La mayoría de las organizaciones tienen sus propios requisitos sobre la forma de estructurar los informes. Si existe flexibilidad con respecto a la forma de estructurar el informe, y la organización que encarga la evaluación no cuenta con un formato establecido, el equipo de evaluación debe concentrarse en lo que resultará más útil para los principales usuarios de la evaluación. Estructurar el informe alrededor de los criterios de evaluación puede ser de utilidad para el gestor de la evaluación y para el departamento de evaluación porque les permite hacer comparaciones entre evaluaciones, pero puede ser menos útil para los directores de programas, a quienes pueden ir dirigidas muchas de las recomendaciones. Por ejemplo, si están evaluando un programa humanitario multisectorial, posiblemente deseen estructurar al menos una parte del informe por sectores, por ejemplo APS o alojamiento.

Una alternativa es estructurar el informe en función de las preguntas de evaluación. Los evaluadores deben considerar compartir el informe con la organización que encarga la evaluación en una etapa temprana del proceso, por ejemplo con el informe inicial, para cerciorarse de que responderá a las necesidades de los usuarios.



### Ejemplos de cómo estructurar un informe de evaluación

Para un informe estructurado de acuerdo a los criterios del CAD de la OCDE, ver la evaluación de la FICR de su programa de medios de vida tras el terremoto de Haití en 2010 en: [www.alnap.org/resource/8864.aspx](http://www.alnap.org/resource/8864.aspx).

Para un informe estructurado en torno a los grandes temas explorados durante la evaluación, ver la evaluación conjunta del Clúster Global de Logística en: <http://documents.wfp.org/stellent/groups/public/documents/reports/wfp251775.pdf>.

El sistema de ALNAP conocido como “*quality proforma*” (ALNAP, 2005), desarrollado como una metodología para evaluar la calidad de los informes de evaluación humanitarios, ofrece una valiosa guía para escribir un informe de EAH y busca lo siguiente:

- Evidencia de que la evaluación valoró la intervención en función de los estándares internacionales apropiados (por ejemplo, el derecho internacional humanitario y de derechos humanos; el Código de Conducta de la Cruz Roja y las ONG, el Proyecto Esfera);
- Que se sustente en un análisis contextual del área y la población afectada (incluidos los factores históricos, sociales, económicos, políticos y culturales relevantes) y que se nutra de dicho análisis para respaldar el análisis de la intervención; y que contenga un análisis claro de la crisis, incluidos los acontecimientos más importantes (y una cronología, si viene al caso);
- Que se aborden determinados temas transversales, entre ellos la igualdad de género, la incidencia política y la atención a los sectores vulnerables y marginados.



#### **Consejo útil**

Elaboren la estructura de su informe antes de iniciar el trabajo de campo. Esto les ayudará a estructurar la recolección y el análisis de la información de modo tal que se ajuste a los requisitos del informe final, reduciendo con ello el trabajo que entraña producirlo. Esto es particularmente útil para cotejar y analizar la información cualitativa recogida en las entrevistas de campo.

Un informe corto tiene más probabilidades de ser leído que uno largo, pero debe suministrar evidencia suficiente para sustentar los hallazgos, conclusiones y recomendaciones. Cuanto menor sea el número de preguntas de evaluación en los TdR, tanto más fácil será balancear estos requerimientos.<sup>1</sup>



#### **Consejo útil**

Si se espera que el informe de evaluación se ciña a un formato determinado, esto debe constar en los TdR. Las especificaciones pueden incluir la extensión del informe así como detalles de la presentación como tipo de letra, espacios, márgenes, etc. Es posible que algunas organizaciones puedan proveer una plantilla para el informe.

## Estructura del informe

Cuadro 17.2: Estructura del informe de evaluación

Sección	Contenido
Carátula y primeras páginas	<p>Información clave, incluyendo:</p> <ul style="list-style-type: none"> <li>• La intervención que está siendo evaluada</li> <li>• El marco cronológico de la evaluación y la fecha del informe</li> <li>• Las ubicaciones (país, región, etc.) de la evaluación – que pueden incluir mapas</li> <li>• Los nombres de los evaluadores y/o a qué institución pertenecen</li> <li>• El nombre de la organización que encarga la evaluación</li> <li>• Un índice que incluya una relación de los cuadros, gráficos, figuras y anexos</li> <li>• Una relación de siglas</li> </ul>
Resumen ejecutivo	<p>Una sección independiente (generalmente de 1-3 páginas) que incluya:</p> <ul style="list-style-type: none"> <li>• Un resumen de la intervención humanitaria que está siendo evaluada</li> <li>• Los objetivos específicos de la evaluación y los destinatarios previstos</li> <li>• Las metodologías de evaluación</li> <li>• Los hallazgos y conclusiones más importantes, en la secuencia en que aparecen en el informe principal</li> <li>• Las principales recomendaciones</li> </ul>
1. Introducción	<p>Alcance y propósito de la evaluación, destinatarios previstos, composición del equipo y estructura del informe</p> <p>Preguntas de evaluación centrales</p> <p>¿Hubo algunas modificaciones a las preguntas de evaluación propuestas en los TdR?</p> <p>¿Cómo se definieron los alcances?</p>
2. Metodologías	<p>Descripción de las principales metodologías utilizadas, adecuación de las mismas y razones para elegir las</p> <p>Si su evaluación se basa en una teoría del cambio, podrían incluirla aquí o referirse a ella e incluirla en un anexo.</p> <p>Naturaleza y alcances de la participación de la población afectada</p> <p>Restricciones más importantes para llevar a cabo la evaluación (por ejemplo, falta de tiempo, acceso restringido a la población afectada, ausencia de línea de base) y efectos de las mismas</p> <p>Eventuales sesgos en el proceso de evaluación o en el equipo de evaluación y cómo fueron mitigados</p> <p>Estándares internacionales utilizados como referente, por ejemplo, las normas mínimas del Proyecto Esfera y las normas de la herramienta de evaluación LEGS, así como marcos conceptuales utilizados, por ejemplo, el marco de desnutrición o el marco de medios de vida referidos en la <a href="#">Sección 5, Enmarcando su evaluación</a></p>
3. Contexto	<p>Análisis contextual de la crisis a la cual está respondiendo la intervención, por ejemplo, área y población afectada, acontecimientos importantes</p>



Sección	Contenido
4. Principales secciones	<p>Organizados por criterios de evaluación, por preguntas de evaluación u otro marco apropiado para la evaluación y sus usuarios previstos, estos capítulos presentan la evidencia y los hallazgos.</p> <div data-bbox="407 272 1057 400" style="background-color: #f0f0f0; padding: 10px; border: 1px solid #ccc;">  <p><b>Consejo útil</b> Presentar las conclusiones y recomendaciones al final de cada sección acentúa (y ayuda a garantizar) que están ancladas en la evidencia, los hallazgos y las conclusiones pertinentes.</p> </div>
5. Conclusiones	<p>Las conclusiones se desprenden de manera lógica de los hallazgos centrales del informe y los reflejan. Proporcionan una base inequívoca y defendible para los juicios de valor. Proveen apreciaciones sobre la intervención que ha sido evaluada y el propósito de la evaluación.</p>
6. Recomendaciones (ver mayores detalles sobre cómo formular las recomendaciones en la <a href="#">Sección 16</a> )	<p>Deben:</p> <ul style="list-style-type: none"> <li>(a) Ser claras, pertinentes y ejecutables, reflejando cualesquier restricciones para hacerles seguimiento;</li> <li>(b) Desprenderse de las principales conclusiones y reflejar las consultas con las partes interesadas más importantes;</li> <li>(c) Ser presentadas en orden de prioridad, con plazos para ponerlas en práctica y sugerencias de los responsables de su seguimiento;</li> <li>(d) Ser limitadas en número.</li> </ul>
Anexos	<p>Los anexos suelen incluir:</p> <ul style="list-style-type: none"> <li>• TdR</li> <li>• Relación de personas entrevistadas y sitios visitados</li> <li>• Relación de documentos consultados e información secundaria utilizada</li> <li>• Detalles adicionales sobre las metodologías, por ejemplo sobre los instrumentos de recolección de información</li> <li>• Información biográfica de los evaluadores y/o justificación de la composición del equipo</li> <li>• Matriz de evaluación</li> <li>• Cronología del tema o la intervención que está siendo evaluada – esto es indispensable en cualquier evaluación que incluya la oportunidad como un criterio</li> </ul> <p>Otros anexos podrían tratar temas cuya discusión detallada estaría fuera de lugar en el informe principal, o presentar resultados de métodos específicos (por ejemplo, un consolidado de respuestas a una encuesta en línea).</p> <div data-bbox="407 1177 1057 1286" style="background-color: #f0f0f0; padding: 10px; border: 1px solid #ccc;">  <p><b>Consejo útil</b> Utilicen anexos para respaldar los elementos y para detalles que abarrotarían el informe principal.</p> </div> <div data-bbox="407 1302 1057 1445" style="background-color: #f0f0f0; padding: 10px; border: 1px solid #ccc;">  <p><b>Consejo útil</b> Si el informe es circulado sobre todo por vía electrónica, consideren presentar los anexos como un documento aparte para así acortar el informe y darle una apariencia menos intimidante al abrirlo por primera vez.</p> </div>

Fuente: Basado en UNEG (2010) y ALNAP (2005)

## La importancia del resumen ejecutivo

La extensión del resumen ejecutivo puede estar reglamentada por las organizaciones, pero en general se sugiere que estos tengan entre dos y cinco páginas. Recuerden que muchas partes interesadas pueden tener el tiempo (o hacer tiempo) para leer el resumen ejecutivo solamente, de modo tal que no solo debe constituir una buena representación de su evaluación sino que también debe autosostenerse. Quizás convenga incluir al final del resumen ejecutivo un resumen de las principales recomendaciones, por ejemplo en forma de cuadro y categorizadas como “imprescindibles, importantes y deseables”.

El resumen ejecutivo no debe figurar en los primeros borradores del informe, para evitar que los comentarios versen sobre el resumen y no sobre el informe completo.



### Consejo útil

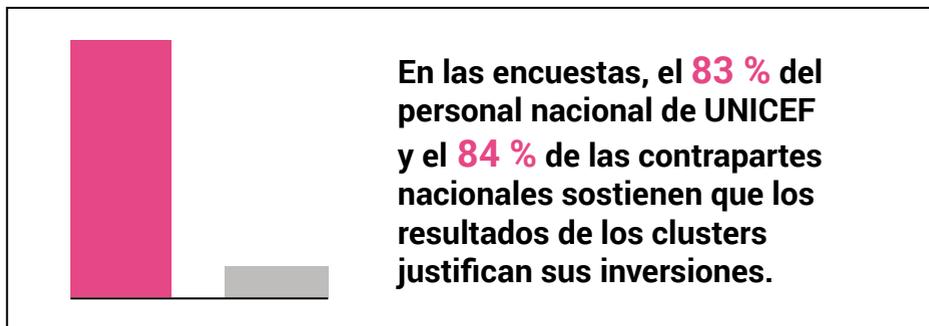
Programen el tiempo para escribir y reescribir el resumen ejecutivo. Esta será la sección más leída del informe, así que no lo dejen para el último minuto, ¡cuando el equipo se esté quedando sin energías!

## Cómo hacer más atractivo su informe

Existen diversas estrategias para hacer del informe un documento más accesible, leíble y atractivo. Ver algunos ejemplos a continuación. Algunas son muy fáciles de implementar, mientras que otras requieren mayor planificación. Es posible que tengan que ser consideradas y convenidas antes del trabajo de campo.

- **Subtítulos:** Utilicen subtítulos claros, que transmitan el mensaje principal de la sección o subsección.
- **Negrita:** La negrita (o cursiva) puede ayudarles a resaltar palabras clave en secciones importantes (pero largas) del informe, por ejemplo, una palabra o frase clave en cada párrafo. Es importante mantener la coherencia en torno al tipo de palabras o frases que se resaltan (por ejemplo, el tema del hallazgo) (DFID, 2005: 43).
- **Citas separadas:** Utilicen citas separadas del texto para destacar afirmaciones particularmente poderosas o importantes.
- **Citas directas:** Las citas directas relacionadas con hallazgos importantes pueden contribuir a animar el informe. Aunque a menudo es vital ocultar el nombre de una persona, el autor de la cita puede ser citado como “informante clave de la organización contraparte”, “maestra de escuela local”, etc. (Sutter et al., 2011). Ver mayores detalles sobre la obtención de consentimiento y el respeto de la confidencialidad en la [Sección 14](#).

- **Recuadros:** Se pueden colocar en recuadros elementos tales como definiciones, miniestudios de caso o ejemplos de buenas prácticas. Estos ayudan a romper la monotonía del informe y a que sobresalgan los mensajes o historias clave.
- **Fotos:** Estas tienen que ser cuidadosamente seleccionadas y contar con los permisos adecuados (derechos de autor o el consentimiento de los individuos que aparecen en la fotografía). Ver la discusión sobre consentimiento informado y confidencialidad en la [Sección 14, Involucrando a la población afectada en su evaluación](#). ¿Cuenta la organización que encarga la evaluación con directrices para indicar qué imágenes pueden y no pueden ser utilizadas?
- **Historias:** Se pueden utilizar relatos cortos para personalizar un hallazgo y ayudar a infundirle emoción al informe de evaluación. Sin embargo, estos deben ser usados con moderación, para apuntalar o servir como ejemplo de un argumento, ya que también tienden a simplificar las cosas. Es importante obtener el consentimiento de las personas aludidas en la historia, especialmente si esta se utiliza fuera del informe o si este es publicado.
- **Cuadros, gráficos y figuras:** Los cuadros, tablas, gráficos y otras figuras pueden comunicar hallazgos clave de manera concisa. Tufté (1990; 1997; 2001) ofrece buenos consejos sobre la presentación de información visual; ver también ejemplos en la [Sección 10](#).
- **Visualización de datos:** El rango de posibles opciones se encuentra en constante evolución y abarca desde infografías estáticas hasta imágenes web que el lector puede explorar. Para indicaciones claras y consejos útiles sobre visualización de datos, ver Evergreen Data (<http://stephanieevergreen.com/blog>).



Evaluación de UNICEF (2013) del rol de UNICEF como organización responsable de clusters en el sector humanitario, documento informativo sobre la evaluación



## Diferentes informes para diferentes destinatarios

Los informes de evaluación con frecuencia tienen un abanico de destinatarios previstos. El gestor de la evaluación y el personal de programas que ejecutó la intervención generalmente desean ver un análisis detallado, y especialmente la evidencia en la cual se sustentan los hallazgos y las conclusiones. Probablemente deseen un informe detallado que explique estos aspectos con detenimiento. Pero es improbable que otros usuarios, por ejemplo los miembros de la alta dirección, lean un informe extenso; por lo tanto, necesitan un resumen de los hallazgos más importantes y sus implicaciones, plasmados en un documento corto. En consecuencia, posiblemente valga la pena producir diferentes versiones del informe final – una versión detallada, que incluya todo, y un resumen corto independiente. DANIDA publica resúmenes informativos cortos junto con cada evaluación. (Ver discusión sobre [canales de comunicación](#) más abajo.)

### Ejemplo de buena práctica: Circular un documento independiente

Un resumen independiente de 16 páginas fue elaborado y difundido para la evaluación conjunta del apoyo a personas desplazadas internamente encargada en 2005 por Sida, los ministerios de asuntos exteriores de Dinamarca y Holanda y ECHO.<sup>2</sup> Esto hizo mucho más accesibles los hallazgos clave del documento completo, de 140 páginas.

Ver Borton et al. (2005)

## Circulación del informe en borrador y comentarios

Los borradores de los informes de evaluación suelen ser circulados por el gestor de la evaluación a las partes interesadas más importantes, entre ellas aquellas que tienen un interés directo en la evaluación y los usuarios previstos de la misma, muchos de los cuales o todos pueden haber sido entrevistados en el marco de la evaluación. Darles la oportunidad de responder a eventuales errores de comprensión o análisis o en los hechos o cifras, es una manera efectiva de asegurar la calidad. Asimismo, son una pieza importante del proceso de promover la apropiación de la evaluación por las partes interesadas. En ocasiones se les da a todos los entrevistados la oportunidad de leer y comentar el borrador del informe, aunque esto rara vez incluye a la población afectada.

## Consejos útiles sobre el proceso

Al solicitar comentarios sobre el informe en borrador.

- Programen un margen de tiempo suficiente para ello – mínimamente entre una y tres semanas. Los receptores del borrador pueden necesitar circularlo dentro de su organización y consolidar posteriormente los comentarios.
- Expliquen sus restricciones de tiempo a las personas que harán los comentarios. Pueden considerar establecer una fecha estricta para la entrega de los comentarios, bajo el supuesto de que la ausencia de respuesta significa aprobación.
- Es útil hacer saber a las partes interesadas el grado de “madurez” del informe. El nivel de profundidad de los comentarios que resulta apropiado para un primer borrador es muy diferente al que se requiere para un borrador final.
- Aclarar el tipo de retroalimentación requerido y cómo debería ser compartido (ver [ejemplo de buena práctica en la pág. 338](#)).
- Si las personas que van a comentar el informe son pocas, posiblemente deseen circularlo en versión Microsoft Word, para que puedan usar las funciones de comentarios y control de cambios, para que cada persona que proporciona retroalimentación sea identificada. Si el informe va a ser comentado por muchas partes interesadas, consideren circular el documento en formato PDF, que no acepta modificaciones, acompañado de una plantilla en Word para comentarios.
- Siempre que sea posible, el gestor de la evaluación deberá consolidar todos los comentarios en un solo documento para que al equipo de evaluación le sea más fácil procesarlos. Esto también contribuye a asegurar que todos los comentarios sobre una sección en particular sean considerados al mismo tiempo, aumentando con ello la eficiencia del proceso de revisión.

Constituye una buena práctica que el equipo de evaluación indique cómo ha respondido a los comentarios de los revisores, incluyendo breves anotaciones sobre los comentarios procesados, por ejemplo:

- Corregido;
- Matizado;
- Detalle añadido;
- Eliminado;
- No aceptado – dando una razón detallada.



### Consejo útil

Numeren los párrafos del informe en borrador para facilitar el proceso de comentar pasajes específicos. No numeren las líneas, sin embargo, ya que esto hace que el proceso de revisión parezca más intimidante y, además, ¡pueden terminar con más comentarios de los que el equipo de evaluación puede manejar!



### Ejemplo de buena práctica: Proceso de retroalimentación en la Cruz Roja Canadiense

La Cruz Roja Canadiense circula los informes de evaluación internamente para recabar comentarios y verificar los hechos y cifras antes de compartirlos más ampliamente con todas las partes interesadas. En la evaluación del despliegue de la Unidad de Respuesta ante Emergencias de Salud tras el paso del tifón Yolanda en Filipinas, el borrador del informe de evaluación fue compartido con todos los entrevistados o encuestados en la fase inicial para que hicieran sus comentarios.

La retroalimentación fue bastante útil, y las partes interesadas apreciaron la oportunidad de revisar los hallazgos. Esto dio un impulso al proceso de retroalimentación y permitió que el informe fuera divulgado antes, mientras el interés en leerlo permanecía alto.

Para recabar la retroalimentación, la coordinadora de la evaluación pidió que los comentarios se organizaran en tres categorías:

- Correcciones y errores: modificaciones factuales
- Diferencias de opinión o información adicional
- Recomendaciones: reflexiones sobre lo que se está proponiendo

Ofreció a las partes interesadas la oportunidad de enviar sus comentarios por correo electrónico o por teléfono para discutir sus ideas.

## Aprobación y finalización del informe de evaluación

Una vez que el equipo de evaluación haya respondido a todos los comentarios recibidos, deberá someter el borrador final del informe a la aprobación ya sea del gestor de la evaluación o de un grupo de gestión más amplio. Hay una serie de documentos referenciales provechosos que pueden ser usados por la organización que encarga la evaluación para la revisión final de la calidad, por ejemplo el sistema *Quality Proforma* de ALNAP, especialmente desarrollado para EAH (ALNAP, 2005), y la guía de orientación de UNEG, que contiene una lista de verificación de la calidad para informes de evaluación (UNEG, 2010).

En esencia, el informe de evaluación debe ser evaluado con base en las siguientes preguntas:<sup>3</sup>

- ¿Responde el informe adecuadamente las preguntas de evaluación? Si la respuesta es negativa, ¿están claramente explicadas las razones para no hacerlo? ¿Son aceptables?
- ¿Proporciona evidencia para sustentar los hallazgos?
- ¿Se desprenden las conclusiones de manera lógica de los hallazgos centrales del informe y los reflejan?
- ¿Las recomendaciones son pertinentes para el propósito de la evaluación, y están respaldadas por evidencia y conclusiones? Ver la [Sección 16, Análisis](#), para mayores detalles sobre la obtención de evidencia para las recomendaciones
- ¿Es coherente el informe y está libre de contradicciones internas?
- ¿Es accesible a sus lectores y usuarios previstos, por ejemplo, en términos de idioma, concisión y diagramación?
- ¿Ha seguido algún requisito de formateo y diagramación?

## 17.3 Difusión<sup>4</sup>

### Planear la difusión desde un inicio

Con demasiada frecuencia se deja la planificación de la difusión de los hallazgos y recomendaciones de la evaluación para después del informe de evaluación. Sin embargo, si se toman en serio la utilización, deben planificar y presupuestar toda la evaluación – incluida la difusión – teniendo en cuenta la difusión (O’Neil, 2012). Esto va mucho más allá de publicar el informe.

Como explican Bamberger et al. (2012: 166), muchas evaluaciones potencialmente valiosas tienen escaso impacto porque los hallazgos no son comunicados a los potenciales usuarios de una forma que les resulte útil o entendible – o, peor aún, porque los hallazgos nunca jamás llegaron a importantes sectores de la comunidad de usuarios.



#### **Ejemplo de buena práctica: Definir la relación de destinatarios de la difusión en los TdR**

El marco de evaluación de la FICR estipula que los TdR deben incluir una relación inicial de los destinatarios de la difusión del informe, “con el fin de asegurar que el informe o resumen de la evaluación llegue a sus destinatarios previstos... La difusión del informe de evaluación puede adoptar una variedad de formas que sean apropiadas para una audiencia específica” (IFRC, 2011a: 15).

He aquí algunas preguntas para ayudarles a planear detalladamente la difusión:

- ¿A qué grupos clave necesitan ser comunicados los hallazgos y recomendaciones de la evaluación?
- ¿Por qué necesita ser comunicada esta información?
- ¿Qué *necesitan* saber estos grupos de destinatarios diferentes?  
¿Qué *les gustaría* saber?
- ¿Cuál es el mejor medio para comunicarse con cada uno de estos grupos?
- ¿Hay algunas consideraciones o limitaciones especiales de los usuarios de la evaluación que debemos tener en cuenta (por ejemplo, conexión de internet irregular, idioma, alta rotación de personal)?
- ¿Cuál es el momento más adecuado para la difusión (por ejemplo, revisión estratégica inminente, nuevo ciclo de planificación)?
- ¿Quién es responsable de la comunicación?

## ¿Informes de evaluación de dominio público?

Algunas organizaciones, generalmente de gran tamaño como Naciones Unidas (por ejemplo, ACNUR), y entidades donantes tienen políticas claras para hacer de dominio público sus informes de evaluación, en aras de la rendición de cuentas. Pero muchas organizaciones no disponen de dichas políticas, prefiriendo decidir en cada caso si deben hacer pública una evaluación o no. Esto puede aumentar el riesgo de que los informes de evaluación más críticos sean “enterrados”, pero en algunos casos los informes no pueden ser publicados por motivos de seguridad, por ejemplo si la publicación de los hallazgos supone un riesgo para el personal o las contrapartes.

Como buena práctica, es importante establecer claramente desde el inicio del proceso de evaluación si el informe será de dominio público o no. De ser necesario, los evaluadores deben pedir una aclaración de este punto cuando sean designados.



### Consejo útil

Las interacciones personales sobre el terreno, así como en las sesiones informativas y los talleres, tienen mayores probabilidades de generar aprendizajes que los informes escritos. Asegúrense de planificar dichas actividades, especialmente en evaluaciones enfocadas en el aprendizaje.

## Diseño de una estrategia de difusión

En su calidad de coordinador de evaluaciones, debería considerar diseñar una estrategia de difusión para cada una de sus evaluaciones, independientemente de cuán amplias o limitadas sean la comunicación y aceptación de los hallazgos de la evaluación. Diseñar una estrategia de difusión significa identificar los canales y productos de comunicación que mejor se ajusten a las necesidades de los diferentes grupos de destinatarios y usuarios de la evaluación.

Algunas recomendaciones pueden tener destinatarios diferentes. En ocasiones se requiere diseñar una estrategia de difusión para cada conclusión y recomendación, para cerciorarse de que lleguen a sus destinatarios previstos y sean usadas adecuadamente. Por ejemplo, las recomendaciones estratégicas y operativas interpelan a niveles muy diferentes de la organización. Es posible que deban considerar metodologías de comunicación distintas para estos dos grupos de destinatarios, focalizadas en lo que cada uno necesita saber.

## Comunicación y retroalimentación a los entrevistados y partes interesadas en el país

Aparte de los talleres informativos al término de la fase de trabajo de campo de una evaluación, a menudo se invierten escasos recursos en la comunicación de los hallazgos a las partes interesadas y entrevistados en el país. Si los usuarios primarios previstos de la evaluación se encuentran en el país, debemos prestarles la debida atención. Consideren presupuestar el retorno del equipo de evaluación a la ubicación o ubicaciones de la evaluación, con el fin de retroalimentar a partes interesadas clave como el personal de programas. Consideren también formas de retroalimentar los hallazgos a los miembros de la población afectada que participaron en la evaluación. Ver el ejemplo de buena práctica de World Vision International (WVI) a continuación.



### **Ejemplo de buena práctica: World Vision International ofrece retroalimentación a la población afectada**

World Vision International utiliza procesos de validación comunitaria para proporcionar retroalimentación a las comunidades consultadas. Esto se lleva a cabo ya sea durante los procesos de diagnóstico, después de las consultas con la comunidad o en el transcurso de las evaluaciones. Estos procesos sirven para informar a las comunidades de la influencia que han tenido las consultas en las decisiones de la organización. El personal de WVI puede elaborar afiches y utilizar papelógrafos u otros materiales para ayudar a presentar los hallazgos a los niños y niñas y a las personas adultas. Otros métodos más creativos utilizados son sesiones de marionetas para niños y niñas.

## 17.4 Diferentes canales de comunicación

Esta sección presenta una serie de canales y productos para socializar los resultados o enseñanzas de una evaluación, ya sea a nivel de la oficina central, en el país concernido o más ampliamente a nivel de la organización. Este listado dista mucho de ser exhaustivo – sean creativos.

### Reuniones, comunicación personal y sesiones informativas

- Pueden organizar reuniones focalizadas entre el equipo de evaluación y los usuarios previstos. Diseñarlas en forma de talleres, y personalizarlas de acuerdo a las necesidades de cada grupo, puede propiciar un mayor compromiso por parte de los usuarios.



#### Consejo útil

¿Podrían ser circuladas más ampliamente las presentaciones resumidas que se utilizan en reuniones y talleres? ¿Podrían ser traducidas? Esto podría facilitar la circulación local de los hallazgos de la evaluación por la organización que encargó la evaluación.

- El equipo de evaluación o el gestor de la evaluación pueden ofrecer sesiones informativas de a dos a usuarios clave.
- Pueden escribir diferentes versiones del informe así como resúmenes o documentos informativos cortos y más accesibles. Como comenta Chris Lysy en su blog de historietas, FreshSpectrum:<sup>5</sup> “No se limiten a escribir un informe de 200 páginas. Escriban un informe de 60 páginas, dos informes de 30 páginas, cinco informes de 10 páginas, 10 informes de cinco páginas y 20 hojas sueltas. Cada uno con un propósito diferente, adaptado a diferentes grupos de destinatarios.” (agosto de 2014)
  - El Programa de Investigación y Políticas de Desarrollo de ODI (RAPID) aplica la regla 1:4:25, es decir, una página para los responsables de las políticas públicas, cuatro páginas para fines informativos y 25 páginas para el informe completo.
  - La evaluación de UNICEF del rol de UNICEF como agencia líder de uno de los cluster de acción humanitaria constituye un buen ejemplo de un resumen de cuatro páginas (UNICEF, 2014).
  - Ver los resúmenes informativos de la Oficina de Evaluación del PMA.<sup>6</sup>



#### Consejo útil

Si no se tiene previsto traducir la versión completa del informe de evaluación a los idiomas indicados, consideren traducir al menos el resumen ejecutivo o un resumen informativo a los idiomas locales para su difusión a las partes interesadas locales.

- Consideren añadir una “carátula” al informe de evaluación, dividiendo las enseñanzas aprendidas en categorías e incluyendo una fotografía instantánea que revele por qué este informe puede ser de interés para quien lo lea (Oliver, 2007: 18).
- Pueden comunicar los hallazgos, conclusiones o recomendaciones de manera sucinta por correo electrónico a usuarios clave.



#### Consejo útil

¿Qué insumos personalizados, basados en la evaluación, podrían hacer llegar a la dirección? Por ejemplo, ¿podrían redactar pautas con los puntos centrales a tratar para la dirección que probablemente tendrá que informar sobre la evaluación en reuniones importantes?

## Material audiovisual

- **Videos:** Los hallazgos y recomendaciones pueden ser comunicados mostrando episodios de la evaluación registrados en video o entrevistas grabadas con el equipo de evaluación. Estos materiales pueden servir como una poderosa herramienta de retroalimentación sobre temas clave comunicados en las palabras de la población afectada, y tenderán a ser más recordados que los largos informes de evaluación. Ver el [ejemplo de buena práctica en la pág. 344](#) y la discusión sobre [consentimiento informado en la Sección 14](#).



#### Consejo útil

El programa PowerPoint permite añadir grabaciones de voz a las diapositivas y convertir en video una presentación. Si no pueden asistir a una reunión, quizás puedan enviar un resumen en video de los hallazgos de la evaluación.

- **Podcasts** (archivos de sonido que pueden escucharse en dispositivos móviles MP3 y en el ordenador) analizando aspectos clave de la evaluación.
- **Webinars** (paneles de discusión o presentaciones en línea), como el webinar que organizó ALNAP conjuntamente con el Comité de Emergencia para Desastres (DEC) sobre el uso de evaluaciones para un cambio real en Filipinas ([www.alnap.org/webinar/19](http://www.alnap.org/webinar/19)) o la serie de webinars sobre evaluaciones de impacto de InterAction ([www.interaction.org/impact-evaluation-notes](http://www.interaction.org/impact-evaluation-notes)).



### Ejemplo de buena práctica: Uso de videos para difundir mensajes o enseñanzas clave

La Oficina de Aprendizaje, Evaluación e Investigación de USAID contrató a un videógrafo para acompañar a un equipo de evaluación en terreno. El cineasta siguió a un equipo de evaluación de proyectos durante un período de tres semanas, con el fin de documentar la mayoría de las fases de la evaluación. Un objetivo era generar aprendizajes acerca del proceso de evaluación, transmitir algunos de los desafíos del trabajo de campo (especialmente en zonas afectadas por conflictos) y proporcionar a los responsables de coordinar, diseñar y ejecutar evaluaciones una mejor apreciación de la complejidad y el valor de la evaluación. El video se encuentra disponible en: [www.usaidlearninglab.org/library/evaluating-growth-equity-mindanao-3-program](http://www.usaidlearninglab.org/library/evaluating-growth-equity-mindanao-3-program).

Dos ejemplos adicionales son el video de Groupe URD sobre la ETR de Haití ([vimeo.com/15198053](https://vimeo.com/15198053)) y el video de OCHA sobre la ETR interinstitucional de la respuesta a las inundaciones de 2011 en Pakistán ([ocha.smugmug.com/Film/OCHAFilms/i-rdN3hsX/A](http://ocha.smugmug.com/Film/OCHAFilms/i-rdN3hsX/A)).

Fuente: Paul Barese en AEA 365 [www.aea365.org/blog/?p=8768](http://www.aea365.org/blog/?p=8768); Linda Morra Imas, intercambio en listserv de IDEAS Evaluation, abril de 2013

### Opciones por internet: medios sociales y blogs

Los medios sociales están evolucionando rápidamente, y constantemente están surgiendo nuevas plataformas y usos. Los medios sociales están siendo aprovechados cada vez más para publicitar la publicación de informes de evaluación o de los hallazgos y recomendaciones de una evaluación. En la actualidad, la mayoría de las plataformas de los medios sociales están orientadas a usuarios externos. Pero existe un amplio espacio para la creatividad aquí.

Una de las reglas básicas para el uso de medios sociales es: aprendan sus reglas o características especiales y adecúen el contenido a ellas. Por ejemplo, publicitar la publicación de un nuevo informe de evaluación a través del Twitter puede ser apropiado, pero sería un post bastante aburrido en un blog o una Comunidad de Prácticas, en cuyo caso publicitar las enseñanzas aprendidas podría ser más atractivo.

He aquí algunos ejemplos:

- **Blogs:** Los detalles sobre los avances de la evaluación del programa de transversalización del enfoque de edad, género y diversidad de ACNUR en Colombia (Mendoza y Thomas, 2009) fueron actualizados en un blog denominado "It Begins with Me. It Begins with You. It Begins with Us" (Empieza conmigo, Empieza contigo. Empieza con nosotros). (Ver <https://itbeginswithme.wordpress.com/>)

- **Historias fotográficas:** Estas podrían ser presentadas en PowerPoint o PDF así como a través de medios sociales como Flickr.com, una plataforma para compartir fotografías que se presta bastante bien a ello. Ver el ejemplo del Banco Asiático de Desarrollo de una historia fotográfica de una evaluación: [www.flickr.com/photos/71616117@N06/sets/72157630309259904/with/7452881380](http://www.flickr.com/photos/71616117@N06/sets/72157630309259904/with/7452881380).
- **Comunidad de Prácticas de Evaluación Humanitaria de ALNAP:** [partnerplatform.org/alnap/humanitarian-evaluation](http://partnerplatform.org/alnap/humanitarian-evaluation).
- **Grupos en LinkedIn:** La Sociedad Europea de Evaluación organiza discusiones en LinkedIn.
- **Twitter:** Groupe URD (@GroupeURD), por ejemplo, es muy activo colgando tuits sobre su trabajo en evaluación.
- **Reddit:** Comparable a un tablero de noticias en línea, Reddit permite a los usuarios colgar vínculos que otros usuarios califican, creando así un *ranking* virtual. El contenido está organizado por temas, llamados *subreddits*. Encontrarán uno sobre evaluación aquí: [www.reddit.com/r/Evaluation](http://www.reddit.com/r/Evaluation).

## Bases de datos de evaluación

Si la gente no puede encontrar su informe, no podrá usarlo. Muchas organizaciones están invirtiendo en mejorar sus bases de datos sobre evaluación. Por ejemplo, la FICR ha rehecho su base de datos en línea (ver [www.ifrc.org/fr/publications/evaluations](http://www.ifrc.org/fr/publications/evaluations)), y Tearfund ha desarrollado una herramienta en Google Docs para compartir a nivel interno los detalles organizativos de las evaluaciones programadas y finalizadas (Hallam y Bonino, 2013; Warner, 2014).

Creada a fines de la década del 90, la biblioteca de evaluación<sup>7</sup> de ALNAP ofrece la colección de materiales de evaluación humanitaria más completa hasta la fecha – tanto informes como metodologías de evaluación y material de orientación, así como artículos seleccionados de investigaciones sobre evaluación. Asegúrense de subir sus productos de evaluación a esta biblioteca.<sup>8</sup>



### Consejo útil

Incluyan los anexos en la versión en línea del informe de evaluación, ya que también existe la posibilidad de colgar una versión sin anexos. Con esto se evitará el problema de que alguien pueda localizar el informe de aquí a cinco años, pero que no pueda encontrar los anexos.

## 17.5 Cómo facilitar la aceptación de una evaluación

Una serie de factores influyen en el uso de los hallazgos y resultados de una evaluación. Ver la [Sección 3](#), así como Hallam y Bonino (2013), para mayores detalles. Por ejemplo, ¿existe una demanda interna de evaluaciones? Promover una demanda interna y sostenerla constituye un proceso de alto nivel y a largo plazo, en el cual una evaluación individual es una única contribución.

Tal como se discutió anteriormente, involucrar a las partes interesadas primarias a todo lo largo del proceso de evaluación es fundamental para el uso y la aceptación de los hallazgos (ver la [Sección 4](#)), al igual que lo es una fase de difusión debidamente planificada y financiada (ver la [Sección 17.4](#)).

La aceptación también puede ser facilitada de varias maneras:

- Involucrando a usuarios clave en la formulación de las recomendaciones. Esto puede acrecentar la apropiación de las recomendaciones por parte de las partes interesadas.
- Identificando personas clave dentro de la organización, comprometidas a actuar y a generar cambios.
- Convirtiendo los hallazgos de evaluaciones individuales (o de varias evaluaciones sobre un mismo tema) en materiales de capacitación, por ejemplo como estudios de caso.
- Asignando claramente responsabilidades de seguimiento. En las organizaciones de mayor tamaño, esto suele hacerse mediante una “matriz de respuesta formal de la dirección”.



### Ejemplo de buena práctica: Desarrollo o mejora de las recomendaciones en consulta con las partes interesadas

En la evaluación de Yogyakarta no se elaboró un borrador de las recomendaciones. En vez de ello, las conclusiones se analizaron en un evento de múltiples partes interesadas en Yogyakarta, con participación de los beneficiarios, representantes del gobierno, ONG nacionales e internacionales y personal de las cuatro organizaciones evaluadas. Las partes interesadas revisaron y corrigieron las conclusiones e hicieron algunas recomendaciones para las ONG internacionales de cara a respuestas futuras. (Wilson et al., 2007: 2-3)

En el caso de Pakistán, el equipo de evaluación se dirigió a las partes interesadas con un borrador de las recomendaciones, pidiéndoles que las mejoraran. Una vez que el informe en borrador estuvo listo, el equipo retornó a Pakistán para realizar tres talleres a nivel provincial y un taller a nivel nacional con partes interesadas clave involucradas en la respuesta humanitaria a las inundaciones.



El jefe del equipo presentó los hallazgos y las conclusiones y recomendaciones en los talleres. Acto seguido, las partes interesadas validaron y priorizaron colectivamente las recomendaciones y decidieron quién sería responsable de poner en práctica cada recomendación así como la línea de tiempo para implementarlas. Los principales cambios en la formulación de las recomendaciones provinieron de las discusiones grupales. El equipo de evaluación consideró que este proceso había incrementado la apropiación de las recomendaciones de la evaluación e impulsado el aprendizaje en tiempo real entre las partes interesadas involucradas (Polastro et al., 2011: 4).

World Vision Internacional también organiza talleres para desarrollar las recomendaciones de algunas evaluaciones. Reúnen a un grupo grande de partes interesadas, en ocasiones de hasta 100 personas repartidas en equipos, en un taller de interpretación de hasta dos días de duración. Los participantes del taller podrían incluir personal senior de dirección, personal sectorial, contrapartes del proyecto, etc. Durante el proceso se entrega a cada grupo secciones diferentes de los hallazgos de la evaluación, en vez del informe completo. Las secciones pueden estar organizadas en torno a los criterios del CAD de la OCDE, por ejemplo. A continuación, cada grupo tiene que trabajar en su sección del informe para responder una serie de preguntas. Este es un trampolín para que los participantes entiendan cómo interpretar los hallazgos y luego los procesen adecuadamente durante el resto del taller, con el objetivo final de generar recomendaciones. De esta manera, los participantes adquieren confianza en el uso de los hallazgos y es mucho más probable que recurran al informe como fuente de información y para tomar decisiones en el futuro.

(Comunicación personal: K. Duryee, World Vision Internacional, 2013; Chamberlain, Jensen y Cascioli Sharp (2014); Vallet (2014))



**Definición: Matriz de respuesta de la dirección**

Registro de la respuesta de la dirección a cada una de las recomendaciones de una evaluación, y medidas que la dirección planea tomar para ponerlas en práctica, con una fecha objetivo y alguien responsable de cada medida.

Figura 17.3: Ejemplo de una matriz de respuesta de la dirección

	¿Financiamiento adicional requerido?		Respuesta de la dirección			Comentario	Acción a tomar	Fecha	Unidad responsable
	Sí	No	Acepta	Acepta parcialmente	Rechaza				
Recomendación 1									
Recomendación 2									

Fuente: FAO

El valor de la matriz de respuesta de la dirección radica en que invita a los usuarios previstos a discutir las recomendaciones y decidir si serán implementadas o cómo. Puede servir también a modo de herramienta de rendición de cuentas (UNEG, 2010). Algunas organizaciones, como el PMA y la FAO, tienen un proceso establecido para hacer un seguimiento de la matriz de respuesta de la dirección, quizás al cabo de seis meses o un año, con el fin de monitorear si la acción propuesta se puso en práctica efectivamente.

A continuación presentamos algunas prácticas recomendables para las respuestas de la dirección, perfiladas por UNEG (2010: 7-8). Estas rigen para evaluaciones tanto grandes como pequeñas.

- **Definir claramente los roles y responsabilidades**, los mismos que deberán ser comunicados a todas las partes interesadas importantes de la evaluación;
- **Establecer un plazo límite acordado**, al término del cual la dirección u otras partes interesadas clave deberán presentar su respuesta formal a la evaluación;
- La dirección debe **designar un punto focal** responsable de coordinar la respuesta de la dirección;
- La respuesta de la dirección debe indicar claramente si **la dirección acepta, acepta parcialmente o rechaza las recomendaciones**. Si se selecciona esta última opción, deberá indicarse la razón o razones.  
En los dos primeros casos, es preciso describir detalladamente las acciones a ser adoptadas, indicando el plazo y la unidad o unidades responsables de ponerlas en práctica. Si se menciona más de una unidad, debe especificarse claramente quién es responsable de qué acción o acciones;
- Las respuestas de la dirección deben ser **reveladas conjuntamente con el informe de evaluación**.

Una limitación de la matriz de respuesta de la dirección se da cuando son varias las organizaciones que han participado en una evaluación conjunta. Pero una posibilidad sigue siendo que una sola organización identifique las recomendaciones que le conciernen y elabore su propia matriz de respuesta de la dirección. En sus buenas prácticas recomendadas, UNEG (2010) sugiere formar un grupo ad hoc de directores que representen a las diferentes organizaciones y/o contrapartes involucradas en una evaluación conjunta para generar una respuesta concertada de la dirección.



**Ejemplo de buena práctica: Involucrando a partes interesadas clave a lo largo de toda la evaluación**

Para la evaluación del Proyecto de Fortalecimiento de las Capacidades de Emergencia de la respuesta al terremoto de Yogyakarta (Wilson et al., 2007), el equipo de evaluación presentó un resumen de los hallazgos al comité de coordinación y al personal, para que estas partes interesadas primarias pudiera trabajar con el equipo en la elaboración de las conclusiones y recomendaciones. Se organizó una reunión final entre el equipo de evaluación, miembros del comité de coordinación, funcionarios del gobierno y la población local con el objeto de revisar y corregir las conclusiones preliminares y formular recomendaciones adicionales.

## 17.6 Síntesis de evaluaciones, revisiones temáticas y metaevaluaciones para potenciar la aceptación

Como explican Hallam y Bonino (2013: 78-80):

La mayoría de los potenciales usuarios de los resultados de una evaluación desean saber algo más que simplemente lo que descubrió una evaluación o investigación individual. “Desean saber qué peso tiene la evidencia” (Weiss, 1998: 317). Decenas de estudios, evaluaciones y revisiones pueden incidir en un mismo problema o tema. Examinarlos en su totalidad, a través de una metaevaluación o una síntesis de evaluaciones, puede arrojar evidencia y hallazgos mucho más sustanciales (Banco Mundial, 2009: 70).



**Definición: Síntesis de evaluaciones**

Una síntesis de evaluaciones es un procedimiento sistemático de organización y resumen de los hallazgos de un conjunto de evaluaciones (posiblemente bastante dispares). Puede constituir una oportunidad para hacer preguntas específicas sobre el conjunto de evaluaciones y/o identificar tendencias comunes en los hallazgos de las evaluaciones.

**Definición: Revisión temática**

Una revisión temática recopila los hallazgos y aprendizajes de una serie de evaluaciones de intervenciones humanitarias pertenecientes a un subsector en particular, por ejemplo, protección o alojamiento. Las revisiones temáticas pueden encargarse para responder y/o afrontar un conjunto específico de preguntas centrales.

**Definición: Metaevaluación**

Una metaevaluación es una evaluación central diseñada para consolidar los hallazgos de una serie de evaluaciones, generalmente en función de un conjunto de parámetros de referencia. Pueden usarse también como “evaluación de una evaluación”, es decir, para evaluar la calidad de un conjunto de evaluaciones.

Existen varios documentos de orientación valiosos que inciden en procesos de síntesis. Aunque es posible que no se refieran a la EAH específicamente, los principios siguen siendo válidos. Por ejemplo, la nota de orientación de Oxfam GB y el Feinstein International Center (sin fecha) para su Programa de Evidencia Humanitaria intenta sintetizar la investigación en el sector humanitario.

Existen varios ejemplos de organizaciones que han reunido y sintetizado el aprendizaje extraído de un abanico de evaluaciones de acciones humanitarias:

- CARE ha realizado una metarevisión de las evaluaciones y exámenes retrospectivos de 15 respuestas a situaciones de emergencia, extrayendo enseñanzas centrales a raíz de este proceso (Oliver, 2007: Anexo 1A).
- El informe anual de evaluación de la Oficina de Evaluación del PMA contiene recomendaciones estratégicas para el PMA, basadas en una síntesis de los hallazgos, conclusiones, enseñanzas y recomendaciones de todas las evaluaciones realizadas por la Oficina de Evaluación el año anterior, e informa sobre actividades de evaluación relacionadas para fortalecer la función de evaluación en el PMA y en el sistema internacional: [www.wfp.org/content/annual-evaluation-report-2013-0](http://www.wfp.org/content/annual-evaluation-report-2013-0).
- NORAD produce periódicamente un informe de síntesis de las enseñanzas extraídas de sus evaluaciones (Disch et al., 2008; Stokke, 2007).
- ACF y Tearfund producen informes de aprendizaje anuales fuertemente basados en los hallazgos de las evaluaciones. Estos son: la Revisión de Aprendizaje de ACF (ACF, 2014) y el Informe de Aprendizaje e Impactos de Tearfund (Tearfund, 2015), respectivamente.

- Desde el 2013, los organismos de las Naciones Unidas llevan a cabo metaevaluaciones anuales con el propósito de evaluar cuán adecuadamente ha sido integrado el enfoque de género en sus informes de evaluación, utilizando un documento proforma desarrollado por UNEG. Dichas metaevaluaciones se utilizan para fines de rendición de cuentas, es decir, para evaluar si las organizaciones están cumpliendo con requerimientos mínimos específicos, basados en las Normas Mínimas y Estándares de Evaluación de UNEG. A principios de 2016, por ejemplo, el PMA contrató consultores independientes para llevar a cabo una metaevaluación de este tipo. En el caso del PMA, la mayoría de sus evaluaciones tienen que ver con el rubro de asistencia humanitaria.

Algunas metarevisiones han explorado hasta qué punto se han tomado en cuenta los hallazgos y recomendaciones de evaluaciones pasadas (Oliver, 2009). Estas pueden enfocarse en temas en los que parece haber una ausencia sostenida de aprendizaje, y pueden explorar provechosamente las razones de la resistencia al cambio y proponer formas de romper con dichos bloqueos en el futuro.

Los documentos de enseñanzas de ALNAP tienen por objeto difundir las enseñanzas de respuestas previas en un formato conciso y legible, con la finalidad de informar y por extensión mejorar el desempeño de acciones humanitarias futuras. Están dirigidos al personal responsable de diseñar y ejecutar las respuestas humanitarias. Los documentos se sustentan principalmente en los hallazgos de evaluaciones almacenados en el Portal de Evaluación y Aprendizaje Humanitarios de ALNAP (HELP, por sus siglas en inglés), así como en otros materiales. Ver todos los documentos de enseñanzas de ALNAP aquí: [www.alnap.org/what-we-do/lessons](http://www.alnap.org/what-we-do/lessons).

## Notas finales

1. El contenido del informe descrito aquí se basa en el sistema *Quality Proforma* de ALNAP (ALNAP, 2005).
2. Ver: [www.sida.se/contentassets/ea2b7ee9cdf40a7a0a54022ebf1167d/support-to-internally-displaced-persons-learning-from-evaluation.-synthesis-report-of-a-joint-evaluation-programme-summary-v\\_3325.pdf](http://www.sida.se/contentassets/ea2b7ee9cdf40a7a0a54022ebf1167d/support-to-internally-displaced-persons-learning-from-evaluation.-synthesis-report-of-a-joint-evaluation-programme-summary-v_3325.pdf).
3. Basado en UNEG (2010) y ALNAP (2005)
4. Esta sección está muy basada en Hallam y Bonino (2013), Área de Capacidad 3, y Warner (2014).
5. Ver: [freshspectrum.com](http://freshspectrum.com).
6. Ver: [www.wfp.org/evaluation/lessons/evaluation-briefs](http://www.wfp.org/evaluation/lessons/evaluation-briefs).
7. Ver [www.alnap.org/resources/results.aspx?type=22](http://www.alnap.org/resources/results.aspx?type=22).
8. [www.alnap.org/account/submitresource.aspx](http://www.alnap.org/account/submitresource.aspx).

## Notas

## Notas

# **Evaluaciones del impacto humanitario**



# 18 / Evaluaciones del impacto humanitario

## 18.1 ¿Por qué es importante la evaluación de impacto?

La mayoría de evaluaciones de intervenciones humanitarias se concentran en los efectos, a menudo de una manera descriptiva. En otras palabras, describen si un efecto deseado ha sido logrado, por ejemplo si las tasas de desnutrición han disminuido, pero tienen que esforzarse por establecer una relación causal entre el efecto y el programa o intervención evaluado.<sup>1</sup> Menos probable aún es que recojan el impacto en términos de los efectos más amplios de una acción humanitaria, pese a que los TdR de la mayoría de las evaluaciones de acciones humanitarias incluyen preguntas relativas al impacto (Proudlock y Ramalingam, 2009; Knox-Clarke y Darcy, 2014: 40-44). Aún existen pocos ejemplos de evaluaciones de intervenciones humanitarias focalizadas exclusivamente en el impacto. Las razones son múltiples, y algunas de ellas se describen más abajo, en la subsección sobre desafíos.

El creciente interés en la evaluación del impacto de las intervenciones humanitarias obedece a que las entidades donantes y organizaciones ejecutoras desean conocer el impacto de las acciones humanitarias que cuentan con financiamiento internacional en las vidas y los medios de sustento de las personas que intentan ayudar. Por ejemplo, las instituciones donantes hacen preguntas acerca del impacto de su – a menudo cuantiosa – inversión en intervenciones humanitarias, generalmente desde la perspectiva de la rendición de cuentas. Las organizaciones ejecutoras pueden querer saber qué respuestas humanitarias son más eficaces, a menudo con fines de aprendizaje.

## 18.2 ¿Qué es el impacto en la EAH?

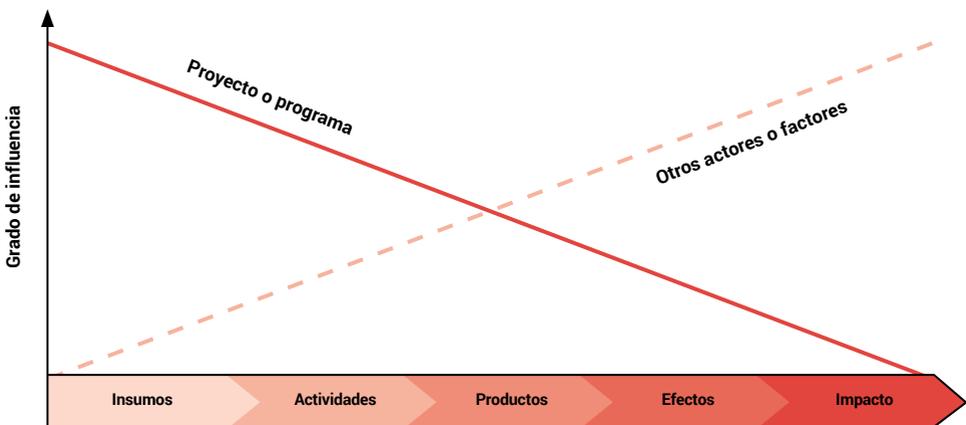
La definición de impacto, de acuerdo a los criterios del CAD de la OCDE, es la siguiente: El impacto se refiere a los efectos más amplios del programa – sociales, económicos, técnicos y medioambientales – sobre los individuos, los grupos categorizados por género y por edad, las comunidades y las instituciones. Los impactos pueden ser previstos o fortuitos, negativos o positivos, a escala macro (sectoriales) o micro (a nivel de hogares, individual), a corto o largo plazo. (Basado en Beck, 2006: 21)

En los hechos, las organizaciones interpretan el impacto de diferentes formas cuando se trata de la evaluación de acciones humanitarias. Algunas se enfocan en los *efectos* que pueden ser atribuidos a una intervención. Por ejemplo, ¿puede atribuirse una reducción de las tasas de desnutrición a la entrega de asistencia alimentaria? Otras organizaciones desean saber acerca del *impacto* más amplio, incluidos los efectos en cadena de la acción humanitaria, intencionales y fortuitos, positivos y negativos. Por ejemplo, ¿hasta qué punto ha afectado la entrega de asistencia alimentaria a los mercados de granos o a la producción alimentaria local?

La esencia de la evaluación de impacto es que explora la causa y el efecto o la "inferencia causal" de las intervenciones. Desplaza el foco de la atención, de la eficacia y la eficiencia de las intervenciones, al examen de si las personas están mejor o más seguras como resultado de ellas. Establecer esta relación causal es el desafío.

En algunos contextos es posible que una evaluación de impacto deba examinar los cambios a largo plazo, perdurables y significativos, que pueden ser atribuidos a una acción humanitaria. Tras un desastre natural, como inundaciones o sequías, por ejemplo, esto podría incluir evidencia de una población mejor preparada o con mayor capacidad de resistencia para enfrentar futuros desastres. En un conflicto prolongado que ha generado un desplazamiento generalizado y a largo plazo, como Sudán del Sur o los movimientos poblacionales asociados a la crisis de Siria, la evaluación puede estar más enfocada en el impacto inmediato, por ejemplo el impacto de la provisión de asistencia humanitaria a las personas desplazadas sobre la población del país anfitrión, en vez del impacto duradero.

Figura 18.1: Influencia relativa a lo largo de la cadena de resultados



Fuente: Smutylo (2001)

El desafío fundamental que entraña la evaluación de impacto es establecer la atribución, en otras palabras, aislar el impacto de una intervención específica de los muchos otros factores que juegan un papel. Como se demostró en la [Figura 18.1](#) (Smutylo, 2001), la influencia de la acción de una organización individual (o incluso de múltiples organizaciones) disminuye en la cadena de resultados a medida que aumenta la influencia de otros factores. Por ejemplo:

- Pese a que las organizaciones humanitarias pueden haber distribuido asistencia alimentaria de manera eficiente y proporcionado atención primaria de salud a la población afectada, el impacto positivo de dichas intervenciones puede haber sido mínimo si el conflicto se intensifica, las comunidades beneficiarias se ven obligadas a huir y la crisis humanitaria se profundiza.
- En sentido inverso, la seguridad alimentaria de una población afectada puede haber mejorado, pero el grado en que este efecto puede ser atribuido a la acción humanitaria es discutible. Otros factores externos pueden haber influido en este resultado; por ejemplo, el incremento de la seguridad alimentaria puede deberse también a una mejora de las condiciones para el funcionamiento de los mercados o la mayor oferta de alimentos disponibles puede responder a factores estacionales como la temporada de cosecha.

El reto de la atribución no es exclusivo de la EAH, pero se ve exacerbado en la mayoría de los contextos humanitarios debido a las dificultades de reunir evidencia de alta calidad. Esto, a su vez, puede deberse a los desafíos contextuales de las crisis humanitarias que se mencionan a continuación, tal como se describe en la [Sección 1, ¿Qué es evaluación de la acción humanitaria?](#):

1. **Desafío:** La volatilidad e inestabilidad del entorno en la mayor parte de los contextos humanitarios, en que una multiplicidad de factores impredecibles afecta los efectos y el impacto, incluida toda la gama de actores diversos.

**Posible solución:** Considerar evaluaciones de impacto conjuntas (ver más abajo).

2. **Desafío:** Falta de información, por ejemplo:
  - a. información básica requerida para diseñar ciertas metodologías de evaluación, por ejemplo ausencia de datos de población o desconocimiento del número de personas afectadas por una crisis;
  - b. información de línea de base sobre indicadores clave para medir el bienestar, por ejemplo medios de vida o acceso a educación, a partir de los cuales se pueda evaluar si se ha generado un cambio;
  - c. información de seguimiento disponible y de alta calidad, que revele los cambios a través del tiempo (la información de seguimiento generalmente se centra en el proceso y los productos, antes que en los efectos).

**Posible solución:** utilizar los recuerdos, y preguntar a los miembros de la población afectada por su situación anterior y de qué manera ésta ha cambiado; utilizar metodologías alternativas para lidiar con la ausencia de información básica, por ejemplo imágenes satelitales para el conteo de las viviendas.

Además de estos desafíos contextuales, algunos desafíos específicos de las evaluaciones afectan la evaluación en el sector humanitario:

1. La necesidad de actuar con rapidez en un entorno impredecible, lo cual significa que hay poco tiempo para preparar una evaluación de impacto por adelantado, es decir, desde las etapas iniciales de la crisis y la respuesta (Puri et al., 2014). Las evaluaciones de impacto tienden a ser planificadas con retraso en el ciclo programático;
2. La selección del diseño más apropiado y la combinación de enfoques más adecuada para responder preguntas específicas sobre la causa y el efecto;
3. La evaluación de impacto requiere niveles de capacidad distintos a los de una evaluación convencional. Los requisitos de recolección y análisis de información pueden hacer necesario un conjunto de capacidades más orientadas a la investigación. Dichas competencias han sido en general escasas en el sector humanitario (Alexander y Bonino, 2015; Proudlock et al., 2009).

Como resultado de estos desafíos, y especialmente de la dificultad de aislar el impacto de una intervención, la evaluación del impacto de la acción humanitaria tiende a enfocarse más en la atribución parcial o en la contribución a un cambio de una intervención en particular (ver la [Sección 1, ¿Qué es evaluación de la acción humanitaria?](#)).

## 18.3 Planificación de una evaluación de impacto

Los desafíos particulares de acometer una evaluación de impacto obligan a asignar una gran cantidad de tiempo a la etapa de planificación. Consideren las siguientes preguntas:

### **¿Están dadas las condiciones para una evaluación de impacto?**

En vista de que las evaluaciones de impacto de alta calidad requieren considerables recursos, asegúrense de que al menos una de las siguientes condiciones se dé para justificar la inversión:

- ¿Es lo suficientemente significativa la intervención (en términos de tamaño, prominencia de la política o posibles consecuencias) como para justificar este tipo de evaluación?

- ¿Es estratégicamente pertinente, en términos del potencial de aprendizaje e influencia?
- ¿Existe evidencia no validada o impugnada de “lo que funciona, para quién y dónde” que la evaluación de impacto propuesta podría ayudar a esclarecer?
- ¿Están dadas las condiciones para que la evaluación sea influyente?  
En otras palabras, ¿cómo se utilizarán los hallazgos? Cerciérense de tener claro este aspecto, para poder enfocarse en la utilización.
- ¿Está preparado el programa o intervención para una evaluación de impacto?  
Por ejemplo, ¿se ha ejecutado durante el tiempo suficiente como para mostrar resultados?
- ¿Se cuenta con un presupuesto suficiente para los tamaños de muestras que se requerirán para demostrar el impacto al tamaño de efecto probable?  
(Ver la [Sección 12, Muestreo](#).)



#### Consejo útil

Una vez que hayan respondido las cinco preguntas arriba indicadas, consideren realizar un análisis de evaluabilidad (ver la [Sección 2, La decisión de realizar una evaluación](#)) antes de embarcarse en una evaluación del impacto humanitario intensiva en recursos.

#### ¿Qué pretenden demostrar o aprender mediante una evaluación de impacto?

Tener claridad sobre este punto ayudará a definir los alcances de la evaluación de impacto. Por ejemplo, ¿la evaluación aspira a establecer si la intervención tuvo en la población afectada el impacto que deseaban los planificadores y financiadores del programa? ¿O está enfocada en el impacto más amplio de la intervención humanitaria, quizás para ubicar su importancia en relación con lo que la población hizo por sí misma?

#### ¿En qué momento del ciclo del programa resulta apropiado realizar una evaluación de impacto?

La respuesta a esta pregunta dependerá en parte de la respuesta a la pregunta anterior. Si les interesa el efecto inmediato de una intervención, quizás deseen iniciar su evaluación de impacto al cabo de unos meses de iniciada la fase de ejecución – por ejemplo, para identificar y entender el impacto inmediato de un programa de alojamiento de emergencia. Si les interesan las consecuencias y el impacto de la acción humanitaria a más largo plazo, quizás deseen realizar su evaluación de impacto varios años después de iniciada la ejecución – por ejemplo, para explorar el impacto de un programa diseñado para fortalecer la capacidad de resiliencia de las comunidades y hogares.

#### ¿Es mejor que sea realizada por una sola organización o en forma de evaluación conjunta?

Las [evaluaciones de impacto conjuntas](#) pueden ayudar a garantizar que se estudie

el impacto global de la acción humanitaria, en vez de intentar aislar artificialmente el trabajo de una sola organización (Beck, 2009: 1). Además, permiten comparar distintos enfoques, los cuales pueden contribuir a evaluar el impacto relativo de diferentes intervenciones. En el sector humanitario, muchas organizaciones que están llevando a cabo evaluaciones de impacto lo están haciendo como un ejercicio conjunto.

#### **¿Cuáles son las preguntas de evaluación apropiadas?**

Una vez que tengan claramente establecido lo que la evaluación de impacto desea averiguar o aprender, será posible poner por escrito las preguntas de evaluación adecuadas. Como se mencionó en la [Sección 6, Selección de las preguntas de evaluación](#), es importante identificar unas cuantas preguntas de alto nivel diseñadas para darles las respuestas que están buscando. Recuerden que sus preguntas de evaluación deben determinar la selección de la metodología y el diseño, y no al revés.

#### **¿Cuáles son el diseño, metodologías y herramientas más idóneos para responder a estas preguntas?**

Para responder a esta pregunta, presten atención a los atributos del programa. Por ejemplo, ¿el programa tiene una teoría del cambio susceptible de ser validada? ¿Es posible crear un contrafactual? Ver la [Sección 18.4, Opuestos](#).

#### **¿Qué tipo de capacidades de evaluación se requieren?**

Una vez que los alcances y las preguntas de evaluación han sido establecidos, y que se tiene cierta noción de las metodologías más apropiadas, el siguiente paso consiste en identificar las capacidades de evaluación necesarias – especialmente para determinar si las capacidades convencionales son suficientes o se requieren evaluadores con capacidades de investigación cualitativa y/o cuantitativa. Proudlock et al. sugieren que “el diseño y ejecución de evaluaciones de impacto requieren capacidades disponibles solamente si se invierte en alianzas a largo plazo entre académicos, donantes, gobiernos, profesionales y beneficiarios focalizados” (2009: 7).

#### **¿Cuáles son las implicaciones presupuestales y en términos de recursos?**

En vista de que las evaluaciones de impacto tienden a requerir un trabajo de campo más largo e intensivo que otros tipos de EAH, y quizás mayores capacidades, es probable que requieran asimismo presupuestos más altos que los de una EAH convencional.



#### **Consejo útil**

Las organizaciones deben priorizar un número reducido de evaluaciones de impacto estratégicas y de alta calidad, en vez de llevar a cabo múltiples evaluaciones de impacto pobremente financiadas y por lo tanto en riesgo de arrojar resultados de baja calidad.

## 18.4 Enfoques y diseños

Los enfoques y metodologías que presentamos a continuación son apropiados para evaluaciones de impacto humanitario porque permiten inferir la causalidad. Ver la Sección 11 para mayores detalles sobre cada una de estas metodologías.

La lista no es exhaustiva y podrían considerarse otros métodos posibles.

En iniciativas lideradas por InterAction y el Centro de Investigaciones Innocenti de UNICEF encontrarán orientación más detallada.<sup>3</sup>

Enfoque / diseño	Potencial	Restricciones	Consejos útiles
<b><u>Pruebas controladas aleatorias o diseño de grupo de comparación</u></b>	Cuando son viables, pueden ayudar a establecer la causalidad, al proveer datos y análisis comparativos creíbles.	Rara vez viables en contextos humanitarios por razones prácticas y éticas. Mediciones iniciales rara vez factibles, solo mediciones finales.	Consideren grupos de comparación oportunistas, por ejemplo campamentos que hayan recibido diferentes niveles o tipos de asistencia.
<b><u>Enfoques basados en teorías</u></b>	Validan la teoría del cambio subyacente de una intervención programática, por ejemplo validando los supuestos.	La teoría del cambio puede no haber sido formulada por los planificadores o ejecutores del programa.	
<b><u>Enfoques basados en casos</u></b>	Pueden usar métodos cualitativos para explicar de qué manera una intervención podría ser responsable de generar ciertos cambios, a partir de un caso específico.	Dificultades para hacer generalizaciones basadas en estudios de caso en contextos diversos y rápidamente cambiantes	Los casos deben ser seleccionados intencionalmente.
<b><u>Estudio longitudinal</u></b>	Realiza una investigación durante un período de tiempo con el fin de identificar los cambios en la vida de la población afectada.	Susceptible de requerir cuantiosos recursos, y es posible que sea inviable en contextos en los que los grupos poblacionales tienen alta movilidad por causa de un conflicto sostenido, o en que el acceso es intermitente.	Es posible que se pueda utilizar información de un estudio longitudinal previo – ver Khoo (2010) para un ejemplo.
<b><u>Enfoque de “contribución al cambio” (Few et al., 2014)</u></b>	Evalúa los cambios a través del tiempo en la vida de la población afectada, el grado al que han recuperado sus medios de vida y su bienestar y el papel que parecen haber cumplido las intervenciones en dicho proceso de recuperación.	Asume que los cambios en el bienestar y los medios de vida de las personas pueden ser identificados con toda claridad a nivel de hogares. Diseñado principalmente para usarse después de desastres de ocurrencia rápida como inundaciones súbitas, tormentas, deslizamientos, terremotos, tsunamis y erupciones volcánicas.	

Un enfoque de metodología mixta resulta particularmente apropiado para la evaluación de impacto humanitario, especialmente cuando los métodos mixtos se sostienen mutuamente al actuar en sinergia (Adato, 2011). Frecuentemente solo las metodologías mixtas son aptas para lidiar con el gran número de factores potencialmente confusos presentes en el contexto típico de la acción humanitaria. El Centro Feinstein ha sido pionero en el uso de un enfoque participativo de evaluación de impacto que crea una asociación de aprendizaje entre la institución donante y las organizaciones ejecutoras (Catley et al., 2009). El ejemplo de buena práctica de la pág. 365 demuestra cómo las metodologías de diagnóstico rural participativo, o diagnóstico rápido participativo (DRP), pueden usarse para explorar el impacto con las comunidades afectadas con un enfoque participativo, incluso en un entorno difícil afectado por conflictos.



#### **Consejo útil**

Al seleccionar el diseño de su evaluación de impacto, consideren qué nivel de certeza y precisión deben tener las respuestas de la evaluación a preguntas específicas para ser considerada verosímil. Esto ayudará a sustentar la decisión sobre el enfoque de muestreo, así como cualquier concesión en términos de presupuesto y metodologías.



#### **Consejo útil**

La consulta extensa y en profundidad con la población afectada juega un papel primordial en la evaluación de impacto. La evaluación de impacto humanitario casi siempre se sostiene en alguna medida en un juicio de valor, lo cual quiere decir que las opiniones de la población afectada deben ser altamente priorizadas para establecer la causalidad.



### **Ejemplo de buena práctica: Serie de evaluaciones conjuntas del PMA y ACNUR sobre el impacto de la asistencia alimentaria en la puesta en práctica de soluciones duraderas para refugiados de larga data**

El PMA y ACNUR encargaron y llevaron a cabo conjuntamente una serie de evaluaciones de impacto basadas en una metodología mixta en Bangladesh, Chad, Etiopía y Ruanda, con el fin de evaluar la contribución de la asistencia alimentaria a la puesta en práctica de soluciones duraderas para el problema de los refugiados de larga data (PMA y ACNUR, 2012). En general, la serie de evaluaciones de impacto empleó un enfoque basado en una teoría como base para inferir la causalidad. Las cuatro evaluaciones utilizaron un mismo marco teórico y un mismo enfoque, pero con detalles adaptados a cada contexto. Algunas de las principales características de esta serie de evaluaciones de impacto son el uso de:

- Un modelo lógico o teoría del cambio desarrollado por primera vez por la Oficina de Evaluación del PMA y analizado y validado posteriormente por los equipos de evaluación en los cuatro países visitados; los equipos evaluaron asimismo la coincidencia entre la TdC y los marcos lógicos nacionales en diferentes etapas de la evaluación.
- Un enfoque basado en una metodología mixta, adaptado a una situación en la que no era posible utilizar un contrafactual convencional como base para inferir la causalidad. El uso de metodologías mixtas para recopilar y analizar la información en los cuatro casos nacionales incluyó: triangulación de la información generada por revisiones documentales; entrevistas con las partes interesadas del PMA y de ACNUR; revisiones de información secundaria; encuestas cuantitativas; caminatas transversales; así como entrevistas cualitativas con grupos focales de beneficiarios y miembros de las comunidades anfitrionas locales, entre otros (PMA y ACNUR, 2012: 2).
- Un análisis evaluativo basado en los resultados que emergieron de los estudios de caso nacionales para establecer: a) qué factores internos y externos podían explicar los resultados en términos de causalidad; y b) qué factores influyeron en los resultados y cambios (intencionales y fortuitos) observados en los diferentes países y por qué (ver PMA y ACNUR, 2012: 10-13).

Fuente: Alexander y Bonino (2015: 11)



### **Ejemplo de buena práctica: Evaluaciones de impacto participativas**

Para un buen ejemplo de una evaluación de impacto participativa, ver la evaluación del programa de la FAO de distribución de kits de apoyo a los medios de vida de emergencia en Sudán del Sur – abril-mayo de 2015.<sup>1</sup>

Como parte de la evaluación del programa de la FAO en Sudán del Sur (focalizada en la respuesta a la emergencia de nivel 3), la Oficina de Evaluación de la FAO efectuó una evaluación de impacto participativa en algunas de las zonas más afectadas por el conflicto en 2014, cuando la FAO había respondido distribuyendo kits de apoyo a los medios de vida con cultivos básicos y semillas de legumbres, herramientas y equipos de pesca.

La evaluación buscaba recoger información de primera mano de las comunidades evaluadas acerca de los impactos positivos, negativos, intencionales y fortuitos de la distribución de dichos materiales, con la finalidad de mejorar las intervenciones de emergencia futuras y escuchar los comentarios de las personas que habían sido afectadas. Se utilizaron metodologías de investigación participativa, especialmente herramientas cualitativas de DRP. Se trazaron mapas de recursos y líneas de tiempo a nivel de aldeas a cargo de grupos focales separados de hombres y mujeres, con el fin de entender los cambios en los activos y vulnerabilidades de las comunidades. Se utilizaron matrices de ingresos, gastos y estrategias de manejo, y se usaron relojes de actividades diarias para evaluar el impacto de la distribución de los kits a nivel de hogares así como aspectos de nutrición. Se visitaron tres áreas geográficas diferentes para incluir ubicaciones bajo el control del gobierno y bajo el control de la oposición.

Se seleccionó un equipo de miembros del personal integrado por personas de origen étnico diverso, lo que les permitió acceder a las diferentes zonas. Estas fueron capacitadas luego en metodologías de DRP y se les pidió llevar a cabo un ensayo de las metodologías en una ubicación “segura”. Acto seguido, desarrollaron un manual de capacitación para adiestrar a encuestadores locales en técnicas de DRP. Los resultados de la evaluación fueron incorporados al programa de emergencia de la FAO en Sudán del Sur en 2016.

Fuente: FAO (copia mimeografiada)

(Ver también el [ejemplo de buena práctica en la pág. 290](#) en la [Sección 15, Acceso restringido](#), para una breve descripción de un estudio de impacto realizado en Somalia como parte de la evaluación efectuada en 2012-2013 de la cooperación de la FAO en Somalia entre 2007 y el 2012.

## Notas finales

1. Ver Alexander y Bonino (2015) y Morra Imas y Rist (2009).
2. Tomadas de Bonino (2015), lo cual se resume en la guía de protección, con base en Rogers (2012) y Chigas, Church y Corlazzoli (2014)
3. InterAction elaboró una Nota sobre la Evaluación de Impacto y una serie de seminarios web que pueden verse en: [www.interaction.org/impact-evaluation-notes](http://www.interaction.org/impact-evaluation-notes).

El Centro de Investigaciones Innocenti de Unicef ha colaborado con la universidad de tecnología y diseño RMIT, BetterEvaluation y la Iniciativa Internacional sobre Evaluación de Impacto (3ie) en la producción de una serie de documentos informativos y videos sobre metodologías de evaluación de impacto: [www.unicef-irc.org/KM/IE/impact.php](http://www.unicef-irc.org/KM/IE/impact.php).

# Glorsario

## **Acción humanitaria / 24**

Los objetivos de la acción humanitaria son: salvar vidas, aliviar el sufrimiento y preservar la dignidad humana tanto durante como después de una crisis y un desastre natural, así como prevenir y fortalecer los esfuerzos de preparación para dichas ocurrencias.

## **Análisis de contenido / 180**

El análisis de contenido es un análisis de información textual que se realiza de manera estandarizada, de modo tal que los evaluadores puedan hacer inferencias con respecto a la información.

## **Análisis de palabras clave / 181**

El análisis de palabras clave es una forma de análisis de contenido que examina la frecuencia con que ocurren ciertas categorías, con la finalidad ya sea de sacar a la luz las tendencias de un mismo documento a través del tiempo o de comparar dos grupos de documentos.

## **Aprendizaje / 27**

Proceso mediante el cual la experiencia y la reflexión generan cambios de conducta o bien la adquisición de nuevas habilidades

## **Atribución / 29**

Imputación de una relación causal entre los cambios observados (o que se espera observar) y una intervención en particular.

## **Codificación / 304**

Codificación significa asignar categorías a piezas de evidencia específicas.

## **Comité de gestión / 150**

Grupo que maneja la evaluación diariamente, entre otras cosas la redacción de los TdR, la contratación y gestión del equipo de evaluación y la gestión de la revisión y culminación del informe de evaluación

## **Comparación antes y después / 204**

Comparación entre la situación de un grupo antes y después de una intervención.

## **Comparación por tratamiento discontinuo / 206**

El diseño de comparación por tratamiento discontinuo compara al grupo justo debajo del punto de corte para recibir asistencia con el grupo justo por encima de éste.

## **Conclusión / 319**

Afirmación inductiva basada en uno o más hallazgos.

## **Contribución / 29**

Analizar la contribución en una evaluación se refiere a buscar formas creíbles de demostrar que una intervención tuvo alguna incidencia en la obtención de los resultados. El análisis de la contribución es una especie de análisis evaluativo que reconoce que el resultado puede deberse a varias causas, incluso si estas, tomadas de manera individual, pueden no ser necesarias o suficientes para generar un impacto.

**Crowd-sourcing / 286**

La técnica de crowd-sourcing o colaboración abierta distribuida utiliza un gran número de voluntarios, ya sea para recolectar información o para analizar la información de imágenes digitales, generalmente por medios indirectos. Este tipo de información se denomina “datos obtenidos por crowd-sourcing”.

**Desviación estándar / 228**

La desviación estándar o desviación típica es una medida de la variabilidad de un parámetro.

**Diagnóstico participativo rápido / 276**

Diagnóstico rápido participativo o diagnóstico rural participativo (DRP) se refiere a un conjunto de metodologías que permiten que la población local enriquezca, analice y comparta su conocimiento y aprendizaje de formas que sean fácilmente entendibles para el equipo de evaluación.

**Diferencias en diferencias / 204**

Este diseño estima los efectos de la asistencia comparando el cambio promedio a través del tiempo en el efecto de interés entre el grupo asistido y un grupo de comparación.

**Difusión / 341**

Proceso de comunicar información a destinatarios específicos con el objeto de ampliar el conocimiento y con miras a modificar políticas y prácticas.

**Diseños cuasi experimentales / 197**

Diseños que utilizan una comparación donde el grupo de comparación no es seleccionado al azar

**Diseño de regresión discontinua / 121**

El diseño de regresión discontinua compara las líneas de regresión de la variable de interés con la puntuación en la que se basó la intervención.

**Diseños experimentales / 199**

Diseños donde las unidades de análisis se asignan aleatoriamente al grupo asistido o al grupo de control. Cada elemento (es decir, cada persona, familia o comunidad) tiene las mismas probabilidades de ser asignado ya sea al grupo asistido o al grupo de control.

**Diseño experimental natural / 208**

Los diseños experimentales naturales realizan una comparación entre un grupo asistido y un grupo similar que no ha sido asistido por azar.

**Diseño longitudinal / 206**

Los estudios longitudinales realizan mediciones repetidas de una misma población a través de los años.

**Diseños no experimentales / 195**

Diseños que prescindan de comparaciones, ya sea entre poblaciones asistidas y no asistidas o de una misma población que ha recibido asistencia a través del tiempo

**Diseño participativo / 202**

La evaluación participativa promueve la participación de todas las partes interesadas en todas las fases de la evaluación, desde la planificación inicial hasta la puesta en práctica de las recomendaciones.

**Efecto de diseño / 225**

El efecto de diseño es el factor por el cual se tiene que modificar el tamaño de la muestra cuando uno se aparta del muestreo aleatorio simple.

**Efectos / 28**

Cambios o giros intencionales o fortuitos en una situación, generados directa o indirectamente por una intervención. Pueden ser deseados (positivos) o no deseados (negativos). Pueden incluir cambios de conducta (actividades, relaciones, políticas, prácticas) en personas individuales, grupos de personas, comunidades, organizaciones, instituciones u otros actores sociales.

**Emparejamiento por puntuación de propensión / 198**

Técnica estadística de coincidencia que intenta emparejar al grupo de comparación con el grupo de control, seleccionando a un grupo que tenga las mismas probabilidades de ser asistido, de acuerdo a las características del grupo

**Estadística descriptiva / 312**

La estadística descriptiva se utiliza para resumir los aspectos más importantes de una población.

**Estadística inferencial / 312**

La estadística inferencial se utiliza para hacer inferencias, ya sea acerca de una población a partir de una muestra o acerca de una hipótesis.

**Estudios de caso / 200**

Descripción y análisis intensivo de uno o más casos (que pueden comprender desde personas individuales hasta Estados) con el fin de extraer conclusiones generales acerca de la intervención.

**Evaluación conjunta / 84**

Evaluación que realizan dos organizaciones o más para evaluar el trabajo de dos organizaciones o más

**Evaluación continua / 83**

Conjunto de evaluaciones diseñadas para llevarse a cabo en el transcurso del tiempo de vida de una intervención

**Evaluación de clusters / 80**

Evaluación de múltiples proyectos pertenecientes a un programa más grande, o bien evaluación vinculada al Sistema de Coordinación de Clusters (o Coordinación Sectorial) de las Naciones Unidas.

**Evaluación de contrapartes / 80**

Evaluación de un conjunto de intervenciones ejecutadas por una sola organización.

**Evaluación de impacto / 81**

Evaluación enfocada en los efectos más generales de un programa humanitario, entre ellos el impacto intencional y fortuito, el impacto positivo y negativo y el impacto a nivel macro (sectorial) y a nivel micro (de hogares, individual)

**Evaluación de la acción humanitaria / 27**

Examen sistemático y objetivo de una intervención humanitaria para determinar el valor o la significación de una actividad, política o programa, con la intención de extraer enseñanzas que permitan mejorar las políticas y prácticas y reforzar la rendición de cuentas.

**Evaluación de políticas / 86**

Evaluación que examina las percepciones, creencias y supuestos que hacen posible, y deseable, un proyecto individual. Pueden evaluar tanto la eficacia de una política en sí como la forma en que dicha política ha sido puesta en práctica.

**Evaluación de proceso / 82**

Evaluación que incide especialmente en los procesos por los cuales los insumos se transforman en productos, y que también puede examinar una intervención en su totalidad

**Evaluación de programas / 79**

Evaluación de un conjunto de intervenciones con un objetivo humanitario unificado.

**Evaluación de proyectos / 79**

Evaluación de una intervención humanitaria individual, con objetivos, recursos y un cronograma de ejecución específicos, que a menudo existe en el contexto de un programa más amplio

**Evaluación de tecnologías / 86**

Evaluación de una técnica o tecnología en especial

**Evaluación de término medio / 83**

Evaluación que se lleva a cabo hacia la mitad de una intervención

**Evaluación de un conjunto de proyectos humanitarios / 81**

Evaluación de toda la cartera de proyectos humanitarios de una organización

**Evaluación en tiempo real / 82**

Evaluación de una operación humanitaria en ejecución mientras se está llevando a cabo

**Evaluación ex-ante / 83**

Evaluación que se lleva a cabo antes de iniciar una intervención

**Evaluación ex-post / 83**

Evaluación que se lleva a cabo al término de una intervención

**Evaluación externa o independiente / 85**

Evaluación efectuada por evaluadores que no pertenecen al equipo ejecutor

**Evaluación institucional / 86**

Evaluación de la dinámica interna de las organizaciones ejecutoras y de sus instrumentos de política, mecanismos de prestación de servicios y prácticas de gestión, así como de los eslabonamientos entre ellos

**Evaluación de una sola agencia / 84**

Evaluación realizada por la organización ejecutora de una intervención

**Evaluación normativa / 82**

Evaluación que compara las acciones ejecutadas con las acciones planeadas o bien con normas específicas

**Evaluación participativa / 85**

Evaluación en la cual las partes interesadas, entre ellas la población afectada, trabajan juntas en el diseño, ejecución e interpretación de los resultados de una evaluación

**Evaluación propia o autoevaluación / 85**

Evaluación realizada por los responsables de diseñar y ejecutar una intervención; también llamada evaluación interna

**Evaluación sectorial / 80**

Evaluación de un grupo de intervenciones pertenecientes a un mismo sector, todas las cuales contribuyen al logro de un objetivo humanitario específico. La evaluación puede abarcar una parte de un país, un país o varios países (UNICEF, 2013).

**Evaluación sistémica / 84**

Evaluación de la respuesta del sistema humanitario internacional a una crisis humanitaria, abierta a todos los actores del sistema

**Evaluación temática / 81**

Evaluación de una selección de intervenciones que inciden en una prioridad humanitaria específica que atraviesa varios países, regiones y posiblemente organizaciones y sectores

**Evaluable / 210**

Sujeto de una evaluación, normalmente un programa o sistema antes que una persona.

**Evidencia / 293**

Evidencia es el cuerpo de hechos o información disponible que indica si una proposición o creencia es verdadera.

**Examen retrospectivo / 265**

Análisis estructurado de una intervención humanitaria, el mismo que permite a un equipo considerar y reflexionar sobre lo que pasó, por qué pasó y cómo sostener las fortalezas y mejorar las debilidades.

**Fase inicial/ 135**

La fase inicial de una evaluación comprende desde la selección del equipo de evaluación hasta la aprobación del informe inicial.

**Grupo de comparación / 205**

Los diseños que utilizan grupos de comparación comparan al grupo asistido con un grupo de comparación seleccionado.

**Grupo de coordinación / 149**

Grupo constituido para encauzar una evaluación a través de etapas centrales como la elaboración de los TdR, la redacción del informe inicial y la redacción del informe final

**Grupo de revisión de pares / 151**

Grupo que presta asesoría sobre aspectos relacionados con la calidad de la evaluación, generalmente formado por evaluadores y otros especialistas seleccionados por su conocimiento de evaluación, de la región o del tipo de intervención bajo evaluación

**Grupo referencial / 150**

Grupo formado por partes interesadas primarias familiarizadas con el entorno local, que pueden prestar asesoría sobre cuestiones prácticas relacionadas con la evaluación y sobre la viabilidad de las recomendaciones

**Hallazgo / 319**

Un hallazgo es una afirmación factual basada en evidencia.

**Impacto / 29**

Se refiere a los efectos más amplios del programa – sociales, económicos, técnicos y ambientales – sobre personas individuales, un género determinado, grupos etarios, comunidades e instituciones. Los impactos pueden ser intencionales y fortuitos, positivos y negativos, a nivel macro (sectorial) y a nivel micro (de hogares, individual), a corto o largo plazo.

**Inferencia causal / 302**

Inferencia causal significa establecer una relación entre una causa y un efecto.

**Información primaria / 293**

Información recolectada para los fines de la evaluación

**Información secundaria / 293**

Información recogida para otros fines pero utilizada por la evaluación

**Instrumento de la encuesta / 233**

Cuestionario que utiliza el entrevistador en las entrevistas formales realizadas en el marco de una encuesta

**Insumos / 28**

Recursos financieros, humanos y materiales empleados en una intervención humanitaria.

**Intervalo de confianza / 227**

Rango dentro del cual esperamos que recaiga el valor de una población en su conjunto.

**Matriz de respuesta de la dirección / 347**

Registro de la respuesta de la dirección a cada una de las recomendaciones de una evaluación y medidas que los directores planean tomar para ponerlas en práctica, con una fecha objetivo y la parte responsable de cada medida.

**Metaevaluación / 87**

Evaluación central diseñada para consolidar los hallazgos de múltiples evaluaciones, generalmente a partir de una serie de parámetros. El término puede utilizarse asimismo para denominar la “evaluación de una evaluación”, es decir, para juzgar la calidad de un grupo de evaluaciones.

**Modelo lógico / 92**

Cuadro o diagrama que revela la teoría programática (la forma en que se espera que los insumos contribuyan al objetivo superior) de una intervención

**Muestreo / 214**

Selección de un subconjunto poblacional para incluirlo en un estudio, en vez de a toda la población

**Muestreo aleatorio / 215**

El muestreo aleatorio extrae una muestra de la población en la que cada miembro de la población tiene la misma posibilidad de ser seleccionado.

**Muestreo estratificado / 138**

Técnica muestral donde primero se divide a la población en segmentos mutuamente excluyentes y luego se toma una muestra aleatoria simple de cada segmento.

**Muestreo intencional / 218**

El muestreo intencional selecciona la muestra de manera intencionada, de modo tal que los elementos muestreados puedan proporcionar la mayor cantidad de información para el estudio.

**Muestreo no aleatorio / 214**

El muestreo no aleatorio elige la muestra en función de alguna característica de la muestra.

**Muestreo por cuotas / 219**

El muestreo por cuotas divide a la población en subcategorías mutuamente excluyentes y luego recoge información para un tamaño o proporción de muestra previamente establecido para cada categoría.

**Muestreo por disponibilidad / 217**

Procedimiento muestral donde la selección se realiza con base en la disponibilidad, la conveniencia del investigador o la autoselección.

**Muestreo por conglomerados / 225**

Muestreo en el cual se toma una muestra de un número de ubicaciones, cada una con un conglomerado de un número de casos determinado.

**Muestreo pseudoaleatorio / 224**

Muestreo donde no se cuenta con un marco muestral. Tradicionalmente, la primera instancia se selecciona aleatoriamente de un punto de partida seleccionado intencionalmente, y las instancias subsiguientes se seleccionan utilizando alguna regla.

**Nivel de confianza / 227**

Probabilidad de que el valor de la población en su totalidad recaiga dentro del intervalo de confianza.

**Poder estadístico / 228**

El poder estadístico es la probabilidad de que un resultado negativo no sea un falso negativo (1 menos el riesgo de un resultado falso negativo).

**Productos / 28**

Productos, bienes y servicios que resultan de una intervención.

**Protección / 25**

Engloba "todas las actividades encaminadas a lograr pleno respeto por los derechos de la persona, en consonancia con la letra y el espíritu de de la normativa pertinente (derechos humanos, derecho humanitario y derecho de los refugiados) (IASC, 2011).

**Prueba controlada aleatorizada / 207**

La prueba controlada aleatorizada (PCA) compara dos grupos seleccionados aleatoriamente, uno de los cuales recibe asistencia, mientras que el otro (el grupo de control) no.

**Saturación de información / 217**

La saturación de información se produce cuando los casos nuevos dejan de aportar conocimiento nuevo.

**Tamaño del efecto / 228**

Diferencia proporcional en la variable de interés entre el grupo tratado y el grupo de control.

**Síntesis de evaluaciones / 349**

Una síntesis de evaluaciones es un procedimiento sistemático de organización y resumen de los hallazgos de un conjunto de evaluaciones (posiblemente bastante dispares entre sí). Pueden constituir una oportunidad para hacer preguntas específicas sobre el conjunto de evaluaciones y/o identificar tendencias comunes en los hallazgos de las evaluaciones.

**Razonamiento evaluativo / 297**

Proceso analítico mediante el cual los evaluadores responden las preguntas de evaluación.

**Recomendación / 312**

Una recomendación es un curso de acción que los evaluadores sugieren como forma de responder a una o más conclusiones.

**Rendición de cuentas / 27**

Medio por el cual el poder se utiliza de manera responsable. Se trata de un proceso mediante el cual se toman en cuenta los puntos de vista de las diferentes partes interesadas y se responde ante ellas, especialmente ante los grupos afectados por la autoridad o el poder.

**Revisión de efectos / 202**

Una revisión de efectos compara los efectos constatados con los efectos programados.

**Revisión de proceso / 201**

Una revisión de proceso compara la forma en que funcionan los procesos con la forma en que se tenía previsto que funcionarían.

**Revisión temática / 350**

Una revisión temática consolida los hallazgos y aprendizajes de una serie de evaluaciones de intervenciones humanitarias pertenecientes a un subsector en particular, por ejemplo, protección o alojamiento. Las revisiones temáticas pueden encargarse para responder y/o afrontar un conjunto específico de preguntas centrales.

**Rúbrica / 178**

Herramienta de calificación utilizada para evaluar un documento en función de una serie de criterios y de manera consistente

**Rúbrica evaluativa / 297**

Cuadro que describe cómo debería lucir la evidencia a distintos niveles de desempeño, en relación con algún criterio de interés o para la intervención en general.

**Seguimiento / 30**

Función sostenida que consiste en recolectar sistemáticamente datos sobre indicadores especificados para proporcionar, a los administradores y a las principales partes interesadas de una intervención humanitaria en curso, indicaciones sobre el grado de avance, el logro de los objetivos y los progresos en el uso de los fondos asignados

**Series temporales interrumpidas / 203**

El diseño de series temporales interrumpidas arroja un estimado del impacto de una intervención después de examinar una serie temporal de datos antes y después de una intervención.

**Teoría del cambio / 97**

Descripción del mecanismo central mediante el cual se produce un cambio para las personas, grupos y comunidades

**Términos de referencia / 118**

Los TdR representan “una descripción general de los requisitos y expectativas de una evaluación. Contienen una declaración explícita de los objetivos de la evaluación, los roles y responsabilidades de los evaluadores y del cliente de la evaluación y los recursos disponibles para la evaluación” (Roberts et al., 2011: 2).

**Uso o utilización / 62**

Una evaluación ha sido utilizada si los usuarios, que tienen la intención y el potencial para llevar a cabo acciones, han prestado atención seria y activamente a sus hallazgos, identificando usos que tengan sentido para sus propios intereses y necesidades (Sandison, 2006: 100-101).

## Bibliografía

Las publicaciones que se nombran a continuación también pueden ser accedidas a través del Portal de Evaluación y Aprendizaje Humanitario (HELP): [www.alnap.org/resources/EHA-2016](http://www.alnap.org/resources/EHA-2016)

**Abate, G. T., de Brauw, A., Minot, N. and Bernard, T. (2014)** *The impact of the use of new technologies on farmers wheat yield in Ethiopia: Evidence from a randomized controlled trial*. Washington, DC: REAP. ([www.alnap.org/resource/22846.aspx](http://www.alnap.org/resource/22846.aspx)).

**Abdullah, M. Y. H., Bakar, N. R., Sulehan, J., Awang, A. H. and Liu, O. P. (2012)** 'Participatory rural appraisal (PRA): An analysis of experience in Darmareja village, Sukabumi District, West Java, Indonesia'. *Akademika*, 82(1): 15–19. ([www.alnap.org/resource/22847.aspx](http://www.alnap.org/resource/22847.aspx)).

**ACAPS. (2011)** *Technical brief: Direct observation and key informant interview techniques for primary data collection during rapid assessments*. Geneva: ACAPS. ([www.alnap.org/resource/22848.aspx](http://www.alnap.org/resource/22848.aspx)).

**ACAPS. (2013)** *How sure are you? Judging quality and usability of data collected during rapid needs assessments*. Geneva: ACAPS. ([www.alnap.org/resource/11437.aspx](http://www.alnap.org/resource/11437.aspx)).

**ACF. (2014)** *ACF Learning Review 2014*. Paris: ACF. ([www.alnap.org/resource/20231.aspx](http://www.alnap.org/resource/20231.aspx)).

**Adato, M. (2011)** *Combining quantitative and qualitative methods for program monitoring and evaluation: Why are mixed-method designs best?* Washington, DC: The World Bank. ([www.alnap.org/resource/8069.aspx](http://www.alnap.org/resource/8069.aspx)).

**Adinolfi, C., Bassiouni, D., Lauritzsen, H. and Williams, H. (2005)** *Humanitarian Response Review*. New York: UN OCHA. ([www.alnap.org/resource/3284.aspx](http://www.alnap.org/resource/3284.aspx)).

**Ager, A., Ager, W., Stavrou, V. and Boothby, N. (2011)** *Inter-agency guide to the evaluation of psychosocial programming in humanitarian crises*. New York: UNICEF. ([www.alnap.org/resource/10214.aspx](http://www.alnap.org/resource/10214.aspx)).

**Ager, A., Stark, S. and Potts, A. (2010)** *Participative ranking methodology: A brief guide: Version 1.1*. New York: Columbia University. ([www.alnap.org/resource/8070.aspx](http://www.alnap.org/resource/8070.aspx)).

**Ager, A., Stark, L., Sparling, T. and Ager, W. (2011)** *Rapid appraisal in humanitarian emergencies using participatory ranking methodology (PRM)*. New York: Columbia University. ([www.alnap.org/resource/22849.aspx](http://www.alnap.org/resource/22849.aspx)).

**Ahmed, L., Malik, Z. and Nawab, B. (2012)** *Humanitarian support to conflict and flood-affected populations in Pakistan*. Paris: ACF. ([www.alnap.org/resource/21080.aspx](http://www.alnap.org/resource/21080.aspx)).

**Alexander, J. (2014)** *Improving accuracy in humanitarian evaluations*. London: ALNAP. ([www.alnap.org/resource/12636.aspx](http://www.alnap.org/resource/12636.aspx)).

**Alexander, J. and Bonino, F. (2014)** *Ensuring quality of evidence generated through participatory evaluation in humanitarian context*. London: ALNAP. ([www.alnap.org/resource/19163.aspx](http://www.alnap.org/resource/19163.aspx)).

**Alexander, J. and Bonino, F. (2015)** *Addressing causation in humanitarian evaluation: A discussion on designs, approaches and examples*. London: ALNAP. ([www.alnap.org/resource/19453.aspx](http://www.alnap.org/resource/19453.aspx)).

**Alexander, J. and Cosgrave, J. (2014b)** *Representative sampling in humanitarian evaluation*. London: ALNAP. ([www.alnap.org/resource/10389.aspx](http://www.alnap.org/resource/10389.aspx)).

**Alexander, J., Darcy, J. and Kiani, M. (2013)** *The 2013 humanitarian accountability report*. Geneva: HAP. ([www.alnap.org/resource/8758.aspx](http://www.alnap.org/resource/8758.aspx)).

**ALNAP. (2002)** *ALNAP annual review 2002: Humanitarian action: improving performance through improved learning*. London: ALNAP. ([www.alnap.org/resource/5189.aspx](http://www.alnap.org/resource/5189.aspx)).

**ALNAP. (2005)** *Assessing the quality of humanitarian evaluations. The ALNAP quality proforma 2005*. London: ALNAP. ([www.alnap.org/resource/5320.aspx](http://www.alnap.org/resource/5320.aspx)).

**ALNAP. (2007a)** *An introduction to evaluation of humanitarian action (EHA) course manual*. London: ALNAP. ([www.alnap.org/resource/22850.aspx](http://www.alnap.org/resource/22850.aspx)).

**ALNAP. (2007b)** *Slow-onset disasters: drought and food and livelihoods insecurity—learning from previous relief and recovery responses*. London: ALNAP. ([www.alnap.org/resource/5243.aspx](http://www.alnap.org/resource/5243.aspx)).

**ALNAP. (2009)** *Participation by crisis-affected populations in humanitarian action. A handbook for practitioners*. London: ALNAP. ([www.alnap.org/resource/5271.aspx](http://www.alnap.org/resource/5271.aspx)).

- ALNAP. (2015)** *The state of the humanitarian system*. London: ALNAP. ([www.alnap.org/resource/21036.aspx](http://www.alnap.org/resource/21036.aspx)).
- ALNAP and DEC. (2013)** *Workshop Summary. Syria Crisis: Evaluators' learning exchange*. London: ALNAP. ([www.alnap.org/resource/9300.aspx](http://www.alnap.org/resource/9300.aspx)).
- ALNAP and Groupe URD. (2009)** *Participation handbook for humanitarian field workers: Involving crisis-affected people in humanitarian response*. London and Plaisians: ALNAP and Groupe URD. ([www.alnap.org/resource/8531.aspx](http://www.alnap.org/resource/8531.aspx)).
- ALNAP, OECD-DAC, and UNEG. (2010)** *Supporting Evaluation in Haiti Mission Report*. London: ALNAP. ([www.alnap.org/resource/22851.aspx](http://www.alnap.org/resource/22851.aspx)).
- Amazon Fund. (2010)** *Logical framework*. Amazon Fund. ([www.alnap.org/resource/22852.aspx](http://www.alnap.org/resource/22852.aspx)).
- AEA. (2003)** *American Evaluation Association response to US Department of Education*. Washington, DC: AEA. ([www.alnap.org/resource/22855.aspx](http://www.alnap.org/resource/22855.aspx)).
- AEA. (2004)** *Guiding principles for evaluators*. Washington, DC: AEA. ([www.alnap.org/resource/22853.aspx](http://www.alnap.org/resource/22853.aspx)).
- Anderson, A.A. (2005)** *The community builder's approach to theory of change: a practical guide to theory development*. New York: The Aspen Institute. ([www.alnap.org/resource/22854.aspx](http://www.alnap.org/resource/22854.aspx)).
- Ashdown, P. (2011)** *Humanitarian emergency response review*. London: DFID. ([www.alnap.org/resource/6355](http://www.alnap.org/resource/6355)).
- Audsley, B., Halme, R. and Balzer, N. (2010)** 'Comparing cash and food transfers: a cost-benefit analysis from rural Malawi', in Were Omamo, S., Gentilini, U. and Sandström, S. (eds) *Revolution: From food aid to food assistance: Innovations in overcoming hunger*. Rome: WFP. ([www.alnap.org/resource/22856.aspx](http://www.alnap.org/resource/22856.aspx)).
- AusAID. (2013)** *Shelter sector response evaluation typhoon Pablo December 2012 in Mindanao, Philippines - shelter cluster report*. Canberra: AusAID. ([www.alnap.org/resource/10074.aspx](http://www.alnap.org/resource/10074.aspx)).
- Baez, J.E. and Santos, I.V. (2007)** *Children's vulnerability to weather shocks: A natural disaster as a natural experiment*. New York: Social Science Research Network. ([www.alnap.org/resource/8078.aspx](http://www.alnap.org/resource/8078.aspx)).

- Bailey, S. (2013)** *Evaluation of Concern Worldwide's emergency response in Masisi, North Kivu, DRC (2012-2013)*. London: Concern. ([www.alnap.org/resource/22857.aspx](http://www.alnap.org/resource/22857.aspx)).
- Baker, J., Chantal, S., Hidalgo, S., Kayungura, G., Posada, S., Tasikasereka, M. and Vinas, M. (2013)** *External evaluation of the rapid response to population movements (RRMP) program in the Democratic Republic of Congo*. Madrid: DARA. ([www.alnap.org/resource/12494.aspx](http://www.alnap.org/resource/12494.aspx)).
- Bamberger, M. (2012)** *Introduction to mixed methods in impact evaluation*. Washington, DC: InterAction. ([www.alnap.org/resource/8079.aspx](http://www.alnap.org/resource/8079.aspx)).
- Bamberger, M., Carden, F. and Rugh, J. (2009)** *Alternatives to the conventional counterfactual: summary of session 713 think tank: American Evaluation Association, Orlando 2009*. Washington, DC: AEA. ([www.alnap.org/resource/8222.aspx](http://www.alnap.org/resource/8222.aspx)).
- Bamberger, M., Rao, V. and Woolcock, M. (2010)** *Using mixed methods in monitoring and evaluation: experiences from international development*. Washington, DC: The World Bank. ([www.alnap.org/resource/22281.aspx](http://www.alnap.org/resource/22281.aspx)).
- Bamberger, M., Rugh, J. and Mabry, L. (2011)** *RealWorld evaluation. Working under budget, time, data, and political constraints: A condensed summary overview. 2nd ed.* London: SAGE. ([www.alnap.org/resource/8076.aspx](http://www.alnap.org/resource/8076.aspx)).
- Bamberger, M. and Segone, M. (2011)** *How to design and manage equity-focused evaluations*. New York: UNICEF. ([www.alnap.org/resource/8080.aspx](http://www.alnap.org/resource/8080.aspx)).
- Barahona, C. (2010)** *Randomised control trials for the impact evaluation of development initiatives: a statistician's point of view*. Rome: ILAC. ([www.alnap.org/resource/22858.aspx](http://www.alnap.org/resource/22858.aspx)).
- Barakat, S. (2006)** *Mid-term evaluation report of the national solidarity programme (NSP), Afghanistan*. York: University of York. ([www.alnap.org/resource/22859.aspx](http://www.alnap.org/resource/22859.aspx)).
- Barakat, S., Hardman, F., Connolly, D., Sundaram, V. and Zyck, S. A. (2010)** *Programme review and evaluability study (PRES): UNICEF's education in emergencies and post-crisis transition (EEPCT) programme*. New York: UNICEF. ([www.alnap.org/resource/8081.aspx](http://www.alnap.org/resource/8081.aspx)).

- Barron, I. G., Abdallah, G. and Smith, P. (2012)** 'Randomized control trial of a CBT trauma recovery program in Palestinian schools'. *Journal of Loss and Trauma*, 2013(18): 306-321. ([www.alnap.org/resource/22860.aspx](http://www.alnap.org/resource/22860.aspx)).
- Baugh, J. B., Hallcom, A. S. and Harris, M. E. (2010)** 'Computer assisted qualitative data analysis software: A practical perspective for applied research'. *Revista del Instituto Internacional de Costos*. ([www.alnap.org/resource/22861.aspx](http://www.alnap.org/resource/22861.aspx)).
- Beall, J. and Schütte, S. (2006)** *Urban livelihoods in Afghanistan*. Kabul: AREU. ([www.alnap.org/resource/22862.aspx](http://www.alnap.org/resource/22862.aspx)).
- Beck, T. (2006)** *Evaluating humanitarian action using the OECD-DAC criteria*. ALNAP Guide. London: ALNAP. ([www.alnap.org/resource/5253.aspx](http://www.alnap.org/resource/5253.aspx)).
- Beck, T. (2009)** *Joint humanitarian impact evaluation: options paper*. London: ALNAP. ([www.alnap.org/resource/5741.aspx](http://www.alnap.org/resource/5741.aspx)).
- Beck, T. (2011)** *Joint humanitarian impact evaluation: report on consultations: Report for the inter-agency working group on joint humanitarian impact evaluation*. London: ALNAP. ([www.alnap.org/resource/6046.aspx](http://www.alnap.org/resource/6046.aspx)).
- Beck, T. and Buchanan-Smith, M. (2008)** 'Joint evaluations coming of age? The quality and future scope of joint evaluations', in *ALNAP 7th Review of Humanitarian Action*. ALNAP Review. London: ALNAP. ([www.alnap.org/resource/5232.aspx](http://www.alnap.org/resource/5232.aspx)).
- Benatar, S. R. (2002)** 'Reflections and recommendations on research ethics in developing countries'. *Social Science & Medicine*, 54(7). ([www.alnap.org/resource/8084.aspx](http://www.alnap.org/resource/8084.aspx)).
- Benham, C., Cascioli Sharp, R., Chamberlain, P., Frederick, J., Gorgonio, R., Jensen, K., Rivera, J. and Veso, Y. (2014)** *Real-time evaluation of World Vision's response to typhoon Haiyan*. Monrovia, CA: World Vision. ([www.alnap.org/resource/20190.aspx](http://www.alnap.org/resource/20190.aspx)).
- Benini, A. (2009)** *Text analysis under time pressure: Tools for humanitarian and development workers*. ([www.alnap.org/resource/8189.aspx](http://www.alnap.org/resource/8189.aspx)).

**Bennett, J., Pantuliano, S., Fenton, W., Vaux, A., Barnett, C. and Brusset, E. (2010)** *Aiding the peace: A multi-donor evaluation of support to conflict prevention and peacebuilding activities in Southern Sudan 2005-2010*. Gatineau: CIDA. ([www.alnap.org/resource/8287.aspx](http://www.alnap.org/resource/8287.aspx)).

**Bentley, M. E., Boot, M., T., Gittelsohn, J. and Stallings, R. Y. (1994)** *The use of structured observations in the study of health behaviour*. The Hague: IRC International Water and Sanitation Centre. ([www.alnap.org/resource/22863.aspx](http://www.alnap.org/resource/22863.aspx)).

**Berenson, M. L., Levine, D. M. and Szabat, K. A. (2013)** 'Estimation and sample size determination for finite populations', in *Basic business statistics: global edition*. Harlow: Pearson Education Limited. ([www.alnap.org/resource/22864.aspx](http://www.alnap.org/resource/22864.aspx)).

**Berger, R., Pat-Horenczyk, R. and Gekkopf, M. (2007)** 'School-based intervention for prevention and treatment of elementary-students' terror-related distress in Israel: A quasi-randomized controlled trial'. *Journal of Traumatic Stress*, 20(4): 541-551. ([www.alnap.org/resource/22865.aspx](http://www.alnap.org/resource/22865.aspx)).

**Berk, R.A. (1983)** 'An introduction to sample selection bias in sociological data'. *American Sociological Review*, 48(3): 386-398. ([www.alnap.org/resource/22866.aspx](http://www.alnap.org/resource/22866.aspx)).

**Berlemann, M., Steinhardt, M. F. and Tutt, J. (2015)** *Do natural disasters stimulate individual saving? Evidence from a natural experiment in a highly developed country*. Bonn: Institute for the Study of Labour. ([www.alnap.org/resource/22867.aspx](http://www.alnap.org/resource/22867.aspx)).

**BetterEvaluation. (2013)** *Understand causes of outcomes and impacts*. Melbourne: BetterEvaluation. ([www.alnap.org/resource/19077.aspx](http://www.alnap.org/resource/19077.aspx)).

**Bhattacharjee, A., Jacob, P. M., Sumasundaram, M., Ramachandran, R., Thileepan, S., Kumar, S., Srikantharajah, S. and Suppiah, L. (2007)** *Final evaluation of CARE Australia supported tsunami response in Trincomalee and Batticaloa districts of Sri Lanka*. Canberra: CARE Australia. ([www.alnap.org/resource/6029.aspx](http://www.alnap.org/resource/6029.aspx)).

**Bhattacharjee, A., Sida, L. and Reddick, M. (2010)** *Evaluation of DFID-Unicef programme of cooperation: Investing in humanitarian action phase III 2006-2009*. New York: UNICEF. ([www.alnap.org/resource/5895.aspx](http://www.alnap.org/resource/5895.aspx)).

**Bhattacharjee, A., Wynter, A., Baker, J., Varghese M. and Lowery, C. (2011)** *Independent review of UNICEF's operational response to the January 2010 earthquake in Haiti*. New York: UNICEF. ([www.alnap.org/resource/6353.aspx](http://www.alnap.org/resource/6353.aspx)).

**Bilukha, O. (2008)** 'Old and new cluster designs in emergency field surveys: in search of a one-fits-all solution'. *Emerging Themes in Epidemiology*, 5(1). ([www.alnap.org/resource/8085.aspx](http://www.alnap.org/resource/8085.aspx)).

**Black, R. E., Allen, L. H., Bhutta, Z. A., Caulfield, L. E., de Onis, M., Ezzati, M., Mathers, C., and Rivera, J. (2008)** 'Maternal and child undernutrition: global and regional exposures and health consequences'. *The Lancet*. ([www.alnap.org/resource/8086.aspx](http://www.alnap.org/resource/8086.aspx)).

**Blake, C. F. (1981)** 'Graffiti and racial insults: the archaeology of ethnic relations in Hawaii', in Gould R. A. and Schiffer, M. B. (eds) *Modern material culture: The archaeology of us*. New York: Academic Press. ([www.alnap.org/resource/22868.aspx](http://www.alnap.org/resource/22868.aspx)).

**Bliss, D. and Campbell, J. (2007)** *Recovering from the Java earthquake: Perceptions of the affected*. San Francisco, CA: Fritz Institute. ([www.alnap.org/resource/8087.aspx](http://www.alnap.org/resource/8087.aspx)).

**Bliss, D. and Larsen, L. (2006)** *Surviving the Pakistan earthquake: Perceptions of the affected one year later*. San Francisco, CA: Fritz Institute. ([www.alnap.org/resource/8088.aspx](http://www.alnap.org/resource/8088.aspx)).

**Boller, K., Buek, K., Burwick, A., Chatterji, M., Paulsell, D., Amin, S., Borkum, E., Campuzano, L., Jacobson, J., Sattar, S., Kadel, S., Pholy, S., Rutajwaha, A., and Sabaa, S. (2011)** *Evaluation of UNICEF's early childhood development programme with focus on government of Netherlands funding (2008-2010): Global synthesis report*. New York: UNICEF. ([www.alnap.org/resource/8089.aspx](http://www.alnap.org/resource/8089.aspx)).

**Bonbright, D. (2012)** *Use of impact evaluation results*. Washington, DC: InterAction. ([www.alnap.org/resource/8534.aspx](http://www.alnap.org/resource/8534.aspx)).

**Bond. (2013)** *Principles and checklist for assessing the quality of evidence*. London: Bond. ([www.alnap.org/resource/22807.aspx](http://www.alnap.org/resource/22807.aspx)).

**Bonino, F. (2014)** *Evaluating protection in humanitarian action : Issues and challenges*. London: ALNAP. ([www.alnap.org/resource/19237.aspx](http://www.alnap.org/resource/19237.aspx)).

**Bonino, F., Jean, I. and Knox Clarke, P. (2014)** *Closing the loop: Effective feedback in humanitarian contexts*. London: ALNAP. ([www.alnap.org/resource/10676.aspx](http://www.alnap.org/resource/10676.aspx)).

**Boonstra, E., Lindbaek, M., Fidzani, B. and Bruusgaard, D. (2001)** 'Cattle eradication and malnutrition in under fives: a natural experiment in Botswana'. *Public Health Nutrition*, 4(4): 877-882. ([www.alnap.org/resource/22869.aspx](http://www.alnap.org/resource/22869.aspx)).

**Boothby, N., Crawford, J. and Halperin, J. (2006)** 'Mozambique child soldier life outcome study: lessons learned in rehabilitation and reintegration efforts'. *Glob Public Health*, 1(1): 87–107. ([www.alnap.org/resource/22870.aspx](http://www.alnap.org/resource/22870.aspx)).

**Borton, J. (ed.) (1994)** *Code of conduct for the International Red Cross and Red Crescent movement and NGOs in disaster relief*. Geneva: IFRC. ([www.alnap.org/resource/8091.aspx](http://www.alnap.org/resource/8091.aspx)).

**Borton, J., Brusset, E. and Hallam, A. (1996)** *The international response to conflict and genocide: Lessons from the Rwanda experience: Humanitarian aid and effects (JEEAR)*. London: ODI. ([www.alnap.org/resource/2517.aspx](http://www.alnap.org/resource/2517.aspx)).

**Borton, J., Millwood, D. (1996)** *The international response to conflict and genocide : Lessons from the Rwanda experience: Study 3: Humanitarian aid and effects*. London: ODI. ([www.alnap.org/resource/9976.aspx](http://www.alnap.org/resource/9976.aspx)).

**Box, G. E. P. and Tiao, G. C. (1975)** 'Intervention analysis with applications to economic and environmental problems'. *Journal of the American Statistical Association*, 70(349): 70-79. ([www.alnap.org/resource/22871.aspx](http://www.alnap.org/resource/22871.aspx)).

**Brancati, D. (2007)** 'Political aftershocks: the impact of earthquakes on intrastate conflict'. *Journal of Conflict Resolution*, 51(5). ([www.alnap.org/resource/8092.aspx](http://www.alnap.org/resource/8092.aspx)).

**Breier, H. (2005)** *Joint evaluations: Recent experiences, lessons learned and options for the future*. Paris: OECD. ([www.alnap.org/resource/8093.aspx](http://www.alnap.org/resource/8093.aspx)).

**British Parachute Association. (2015)** How safe? ([www.alnap.org/resource/22872.aspx](http://www.alnap.org/resource/22872.aspx)).

**Brookhart, M. A. Wyss, R., Layton, J. B. and Stürmer, T. (2013)** 'Propensity score methods for confounding control in nonexperimental research'. *Circulation: Cardiovascular Quality and Outcomes*. ([www.alnap.org/resource/22873.aspx](http://www.alnap.org/resource/22873.aspx)).

- Broughton, B., Maguire, S. and David-Toweh, K. (2006)** *Inter-agency real-time evaluation of the humanitarian response to the Darfur crisis*. New York: OCHA. ([www.alnap.org/resource/3245.aspx](http://www.alnap.org/resource/3245.aspx)).
- Brown, D. and Donini, A. (2014a)** *Rhetoric or reality? Putting affected people at the centre of humanitarian action*. London: ALNAP. ([www.alnap.org/resource/12859.aspx](http://www.alnap.org/resource/12859.aspx)).
- Brown, D., Donini, A., and Knox Clarke, P. (2014b)** *Engagement of crisis-affected people in humanitarian action. ALNAP Background paper*. London: ALNAP. ([www.alnap.org/resource/10439.aspx](http://www.alnap.org/resource/10439.aspx)).
- Brown, M., Neves, J., Sandison, P., Buchanan-Smith, M., and Wiles, P. (2005)** *Evaluation of DFID-UNICEF programme of cooperation to strengthen UNICEF programming as it applies to humanitarian response, 2000-2005*. New York: UNICEF. ([www.alnap.org/resource/3355.aspx](http://www.alnap.org/resource/3355.aspx)).
- Brusset, E., Bhatt, M., Bjornestad, K., Cosgrave, J., Davies, A., Deshmukh, Y., Haleem, J., Hidalgo, S., Immajati, Y., Jayasundere, R., Mattsson, A., Muhaimin, N., Polastro, R. and Wu, T. (2009)** *A ripple in development? Long term perspectives on the response to the Indian Ocean tsunami 2004*. Stockholm: SIDA. ([www.alnap.org/resource/5679.aspx](http://www.alnap.org/resource/5679.aspx)).
- Brusset, E., Pramana, W., Davies, A., Deshmukh, Y. and Pedersen, S. (2006)** *Links between relief, rehabilitation and development in the tsunami response: Indonesia Case Study*. London: TEC. ([www.alnap.org/resource/5421.aspx](http://www.alnap.org/resource/5421.aspx)).
- Bryman, A. (2004)** 'Triangulation', in Lewis-Beck, M. S., Bryman, A. and Liao, T. F. (eds) *Encyclopedia of social science research methods*. Thousand Oaks, CA: SAGE. ([www.alnap.org/resource/22874.aspx](http://www.alnap.org/resource/22874.aspx)).
- Buchanan-Smith, M., Leyland, T., Aklilu, Y., Noor, M., Ahmed, S., Dini, S. and Robinson, I. (2013)** *Evaluation of FAO's cooperation in Somalia: 2007 to 2012*. Rome: FAO. ([www.alnap.org/resource/8820.aspx](http://www.alnap.org/resource/8820.aspx)).
- Buchanan-Smith, M. et al. (2005)** 'How the Sphere Project came into being: A case study of policy making in the humanitarian-aid sector and the relative influence of research', in Court, J., Hovland, I. and Young, J. (eds). *Bridging research and policy in development: Evidence and the change process*. London: ODI. ([www.alnap.org/resource/23454.aspx](http://www.alnap.org/resource/23454.aspx)).

- Buchanan-Smith, M., Ong, J. C. and Routley, S. (2015)** *Who's listening? Accountability to affected people in the Haiyan response*. Woking: Plan International. ([www.alnap.org/resource/20632.aspx](http://www.alnap.org/resource/20632.aspx)).
- Burns, M., Rowland, M., N'Guessan, R., Carneiro, I., Beeche, A., Sesler Ruiz, S., Kamara, S., Takken, W., Carnevale, P. and Allan R. (2012)** 'Insecticide-treated plastic sheeting for emergency malaria prevention and shelter among displaced populations: An observational cohort study in a refugee setting in Sierra Leone'. *The American Journal of Tropical Medicine and Hygiene*, 87(2): 242–250. ([www.alnap.org/resource/22875.aspx](http://www.alnap.org/resource/22875.aspx)).
- Bush, J. and Ati, H.A. (2007)** *Oxfam's cash transfers in the Red Sea state*. Oxford: Oxfam. ([www.alnap.org/resource/8187.aspx](http://www.alnap.org/resource/8187.aspx)).
- Bush, K. and Duggan, C. (2013)** 'Evaluation in conflict zones: Methodological and ethical challenges'. *Journal of Peacebuilding & Development*, 8(2): 5–25. ([www.alnap.org/resource/22876.aspx](http://www.alnap.org/resource/22876.aspx)).
- Buttenheim, A. (2009)** *Impact evaluations in the post-disaster setting: A conceptual discussion in the context of the 2005 Pakistan earthquake*. New Delhi: 3ie. ([www.alnap.org/resource/8206.aspx](http://www.alnap.org/resource/8206.aspx)).
- Caliendo, M. and Kopeinig, S. (2005)** *Some practical guidance for the implementation of propensity score matching*. Bonn: IZA. ([www.alnap.org/resource/22877.aspx](http://www.alnap.org/resource/22877.aspx)).
- Cameron, L. A. and Shah, M. (2012)** *Risk-taking behavior in the wake of natural disasters*. Bonn: IZA. ([www.alnap.org/resource/22878.aspx](http://www.alnap.org/resource/22878.aspx)).
- Campo, R. (2006)** "Anecdotal evidence": why narratives matter to medical practice'. *PLoS Medicine*, 3(10), 1677-1678. ([www.alnap.org/resource/22879.aspx](http://www.alnap.org/resource/22879.aspx)).
- Canteli, C., Morris, T. and Steen, N. (2012)** *Synthesis of mixed method impact evaluations of the contribution of food assistance to durable solutions in protracted refugee situations*. Rome: WFP. ([www.alnap.org/resource/19902.aspx](http://www.alnap.org/resource/19902.aspx)).
- Carden, F. (2009)** *Knowledge to policy: making the most of development research*. New Delhi: SAGE. ([www.alnap.org/resource/22881.aspx](http://www.alnap.org/resource/22881.aspx)).
- Carlsson, J., Eriksson-Baaz, M., Fallenius, A. and Lovgren, E. (1999)** *Are evaluations useful? Cases from Swedish Development co-operation*. Stockholm: SIDA. ([www.alnap.org/resource/12371.aspx](http://www.alnap.org/resource/12371.aspx)).

- Catalano, R., Bruckner, T., Gould, J., Eskenazi, B., and Anderson, E. (2005)** 'Sex ratios in California following the terrorist attacks of September 11, 2001'. *Human Reproduction*, 20(5): 1221–1227. ([www.alnap.org/resource/22882.aspx](http://www.alnap.org/resource/22882.aspx)).
- CRS. (2012)** *CRS Haiti accountability framework*. Baltimore, MD: CRS. ([www.alnap.org/resource/10549.aspx](http://www.alnap.org/resource/10549.aspx)).
- Catley, A., Burns, J., Abebe, D. and Suji, O. (2008)** *Participatory impact assessment: A guide for practitioners*. Somerville, MA: Feinstein International Centre. ([www.alnap.org/resource/8094.aspx](http://www.alnap.org/resource/8094.aspx)).
- Catley, A., Burns, J., Abebe, D. and Suji, O. (2013)** *Participatory impact assessment: A design guide*. Somerville, MA: Feinstein International Centre. ([www.alnap.org/resource/10811.aspx](http://www.alnap.org/resource/10811.aspx)).
- CDC. (2009)** *Data collection methods for program evaluation: interviews*. Atlanta, GA: CDC. ([www.alnap.org/resource/19260.aspx](http://www.alnap.org/resource/19260.aspx)).
- Channel Research. (2011)** *5-year evaluation of the Central Emergency Response Fund*. London: Channel Research. ([www.alnap.org/resource/6357.aspx](http://www.alnap.org/resource/6357.aspx)).
- Chapman, N. and Vaillant, C. (2010)** *Synthesis of country programme evaluations conducted in fragile states*. London: DFID. ([www.alnap.org/resource/6358.aspx](http://www.alnap.org/resource/6358.aspx)).
- Chatham House. (2007)** *Chatham house rules*. London: Chatham House. ([www.alnap.org/resource/8097.aspx](http://www.alnap.org/resource/8097.aspx)).
- Checchi, F., Roberts, B. and Morgan, O. (2009)** *A new method to estimate mortality in crisis-affected populations: Validation and feasibility study*. Washington, DC: USAID. ([www.alnap.org/resource/22883.aspx](http://www.alnap.org/resource/22883.aspx)).
- Christensen, G. (2016)** *Manual of best practices in transparent social science research*. Berkeley, CA: Berkeley Initiative for Transparency in the Social Sciences. ([www.alnap.org/resource/22884.aspx](http://www.alnap.org/resource/22884.aspx)).
- Christoplos, I. (2006)** *Links between relief, rehabilitation and development in the tsunami response*. London: TEC. ([www.alnap.org/resource/3533.aspx](http://www.alnap.org/resource/3533.aspx)).
- CHS Alliance. (2014)** *Core humanitarian standard on quality and accountability*. Geneva: CHS Alliance. ([www.alnap.org/resource/22799.aspx](http://www.alnap.org/resource/22799.aspx)).

- CHS Alliance. (2015)** *Core Humanitarian CHS Guidance Notes and Indicators*. Geneva: CHS Alliance. ([www.alnap.org/resource/22885.aspx](http://www.alnap.org/resource/22885.aspx)).
- CIDA. (2000)** *How to perform evaluations: Model ToR*. Gatineau: CIDA. ([www.alnap.org/resource/22886.aspx](http://www.alnap.org/resource/22886.aspx)).
- CIDA. (2005)** *Additional resources: Accompanying document of the environment handbook for community development initiatives*. Gatineau: CIDA. ([www.alnap.org/resource/8098.aspx](http://www.alnap.org/resource/8098.aspx)).
- Clarke, N., Loveless, J., Ojok, B., Routley, S. and Vaux, T. (2015)** *Report of the inter-agency humanitarian evaluation (IAHE) of the response to conflict in South Sudan*. New York: OCHA. ([www.alnap.org/resource/22828.aspx](http://www.alnap.org/resource/22828.aspx)).
- Collier, D. and Mahoney, J. (1996)** 'Insights and pitfalls: Selection bias in qualitative research'. *World Politics*, 49(1): 56–91. ([www.alnap.org/resource/22887.aspx](http://www.alnap.org/resource/22887.aspx)).
- Collinson, S. and Elhawary, S. (2012)** *Humanitarian space: a review of trends and issues*. London: ODI. ([www.alnap.org/resource/22888.aspx](http://www.alnap.org/resource/22888.aspx)).
- Cook, T.D., Scriven, M., Coryn, C. L. S. and Evergreen, S. D. H. (2010)** 'Contemporary thinking about causation in evaluation: A dialogue with Tom Cook and Michael Scriven'. *American Journal of Evaluation*, 31(1): 105–117. ([www.alnap.org/resource/22323.aspx](http://www.alnap.org/resource/22323.aspx)).
- Corlazzoli, V. and White, J. (2013)** *Practical approaches to theories of change in conflict, security, and justice. Part II: Using theories of change in monitoring and evaluation*. London: DFID. ([www.alnap.org/resource/22890.aspx](http://www.alnap.org/resource/22890.aspx)).
- Cornwall, A. (2014)** *Using participatory process evaluation to understand the dynamics of change in a nutrition education programme*. Brighton: IDS. ([www.alnap.org/resource/22891.aspx](http://www.alnap.org/resource/22891.aspx)).
- Cosgrave, J. (2004)** *Danish assistance to internally displaced persons in Angola 1999-2003*. London: Channel Research. ([www.alnap.org/resource/22892.aspx](http://www.alnap.org/resource/22892.aspx)).
- Cosgrave, J. (2008)** *Responding to earthquakes 2008: Learning from earthquake relief and recovery operations*. London: ALNAP. ([www.alnap.org/resource/5239.aspx](http://www.alnap.org/resource/5239.aspx)).

- Cosgrave, J. (2010)** *Evaluability assessment of a proposed evaluation of humanitarian interventions in South and Central Somalia: Commissioned by the IASC for Somalia: overall report*. Geneva: IASC. ([www.alnap.org/resource/22893.aspx](http://www.alnap.org/resource/22893.aspx)).
- Cosgrave, J., Crawford, N. and Mosel, I. (2015)** *10 Things to know about refugees and displacement*. London: HPG/ODI. ([www.alnap.org/resource/22895.aspx](http://www.alnap.org/resource/22895.aspx)).
- Cosgrave, J., Goncalves, C., Martyris, D., Polastro, R. and Sikumba-Dils, M. (2007)** *Inter-agency real-time evaluation of the response to the February 2007 floods and cyclone in Mozambique*. Madrid: DARA. ([www.alnap.org/resource/3457.aspx](http://www.alnap.org/resource/3457.aspx)).
- Cosgrave, J., Polastro, R. and Zafar, F. (2010)** *Inter-agency real time evaluation (IA RTE) of the humanitarian response to Pakistan's 2009 displacement crisis*. Geneva: IASC. ([www.alnap.org/resource/11896.aspx](http://www.alnap.org/resource/11896.aspx)).
- Cosgrave, J., Ramalingam, B. and Beck, T. (2009)** *Real-time evaluations of humanitarian action*. ALNAP Guide. London: ALNAP. ([www.alnap.org/resource/5595.aspx](http://www.alnap.org/resource/5595.aspx)).
- Cosgrave, J., Selvester, K., Fidalgo, L., Hallam, A. and Taimo, N. (2001)** *Independent evaluation of DEC Mozambique floods appeal funds March 2000 - December 2000 : Volume One: Main Findings*. London: DEC. ([www.alnap.org/resource/2999.aspx](http://www.alnap.org/resource/2999.aspx)).
- Cosgrave, J., Wata, H., Ntata, P., Immajati, Y. And Bhatt, M. (2010)** *Programme evaluation of disaster risk reduction commissioned by Cordaid overall report*. London: Channel Research. ([www.alnap.org/resource/12523.aspx](http://www.alnap.org/resource/12523.aspx)).
- Cossée, O., Belli, L., Bultemeier, B. and Carrugi, C. (2010)** *Evaluation of FAO interventions funded by the CERF: Final report*. Rome: FAO. ([www.alnap.org/resource/8099.aspx](http://www.alnap.org/resource/8099.aspx)).
- Court, J. and Young, J. (2003)** *Bridging research and policy: Insights from 50 case studies*. London: ODI. ([www.alnap.org/resource/8100.aspx](http://www.alnap.org/resource/8100.aspx)).
- CP MERG. (2012)** *Ethical principles, dilemmas and risks in collecting data on violence against children: A review of available literature*. CP MERG. ([www.alnap.org/resource/22327.aspx](http://www.alnap.org/resource/22327.aspx)).

- Crawford, N., Cosgrave, J., Haysom, S., Walicki, N. (2015)** *Protracted displacement: uncertain paths to self-reliance in exile*. London: HPG/ODI. ([www.alnap.org/resource/21304.aspx](http://www.alnap.org/resource/21304.aspx)).
- Crawford, P., Bysouth, K., Nichols, D. and Thompon, F. (2006)** *CAER cluster evaluation: Pakistan earthquake*. Canberra: AusAID. ([www.alnap.org/resource/8190.aspx](http://www.alnap.org/resource/8190.aspx)).
- Crawford, P. and Eagles, J. (2007)** *ANCP Pacific cluster evaluation report*. Version 2.0. Canberra: AusAID. ([www.alnap.org/resource/8191.aspx](http://www.alnap.org/resource/8191.aspx)).
- Creswell, J.W. (2002)** *Educational research: planning, conducting, and evaluating quantitative and qualitative research*. 4th ed. Upper Saddle River: Pearson. ([www.alnap.org/resource/22896.aspx](http://www.alnap.org/resource/22896.aspx)).
- Creswell, J. W. and Plano Clark, V. L. (2011)** *Designing and conducting mixed methods research*. 2nd ed. Thousand Oaks, CA: SAGE. ([www.alnap.org/resource/22897.aspx](http://www.alnap.org/resource/22897.aspx)).
- Crewe, E. and Young, J. (2002)** *Bridging research and policy: Context, evidence and links*. London: ODI. ([www.alnap.org/resource/22898.aspx](http://www.alnap.org/resource/22898.aspx)).
- Crisp, J., Garras, G., McAvoy, J., Schenkenberg, E., Speigel, P. and Voon, F. (2013)** *From slow boil to breaking point: A real-time evaluation of UNHCR's response to the Syrian refugee emergency*. Geneva: UNHCR. ([www.alnap.org/resource/8848.aspx](http://www.alnap.org/resource/8848.aspx)).
- Curtis, V. and Kanki, B. (1999)** *A manual on hygiene promotion*. New York: UNICEF. ([www.alnap.org/resource/22899.aspx](http://www.alnap.org/resource/22899.aspx)).
- Dabelstein, N. (1996)** 'Evaluating the international humanitarian system: rationale, process and management of the joint evaluation of the international response to the Rwanda genocide. *Disasters*, 20(4). ([www.alnap.org/resource/8101.aspx](http://www.alnap.org/resource/8101.aspx)).
- Dahlgren, A.-L., DeRoo, L. A., Avril, J. and Loutan, L. (2009)** 'Health risks and risk-taking behaviors among International Committee of the Red Cross (ICRC) expatriates returning from humanitarian missions. *Journal of Travel Medicine*, 16(6): 382–390. ([www.alnap.org/resource/22900.aspx](http://www.alnap.org/resource/22900.aspx)).

**Daito, H., Suzuki, M., Shiihara, J., Kilgore, P. E., Ohtomo, H., Morimoto, K., Ishida, M., Kamigaki, T., Oshitani, H., Hashizume, M., Endo, W., Hagiwara, K., Ariyoshi, K. and Okinaga, S. (2013)** 'Impact of the Tohoku earthquake and tsunami on pneumonia hospitalisations and mortality among adults in northern Miyagi, Japan: a multicentre observational study'. *Thorax*, 68(6): 544–550. ([www.alnap.org/resource/22901.aspx](http://www.alnap.org/resource/22901.aspx)).

**DANIDA. (2012)** *Evaluation of Danish development support to Afghanistan*. Copenhagen: DANIDA. ([www.alnap.org/resource/22903.aspx](http://www.alnap.org/resource/22903.aspx)).

**Daniel, J., (2013)** *Sampling essentials: Practical guidelines for making sampling choices*. Los Angeles, CA: SAGE. ([www.alnap.org/resource/9161.aspx](http://www.alnap.org/resource/9161.aspx)).

**Das, N. C. and Shams, R. (2011)** *Asset transfer programme for the ultra poor: A randomized control trial evaluation*. Dhaka: BRAC. ([www.alnap.org/resource/22904.aspx](http://www.alnap.org/resource/22904.aspx)).

**Davidson, E. J. (2007)** 'Unlearning some of our social scientist habits'. *Journal of MultiDisciplinary Evaluation*, 4(8): iii–vi. ([www.alnap.org/resource/22906.aspx](http://www.alnap.org/resource/22906.aspx)).

**Davidson, E. J. (2009)** *Causal inference: Nuts and bolts: A mini workshop for the anzea Wellington branch*. Wellington: Davidson Consulting. ([www.alnap.org/resource/22908.aspx](http://www.alnap.org/resource/22908.aspx)).

**Davidson, E. J. (2010a)** '9 golden rules for commissioning a waste-of-money evaluation', in Rogers, P. J. and Davidson, E. J. (eds) *Genuine Evaluation*. ([www.alnap.org/resource/22909.aspx](http://www.alnap.org/resource/22909.aspx)).

**Davidson, E. J. (2010b)** 'Commissioning XGEMs – the sequel', in Rogers, P. J. and Davidson, E. J. (eds) *Genuine Evaluation*. ([www.alnap.org/resource/22910.aspx](http://www.alnap.org/resource/22910.aspx)).

**Davidson, E. J. (2010c)** 'Extreme genuine evaluation makeovers (XGEMs) for commissioning', in Rogers, P. J. and Davidson, E. J. (eds) *Genuine Evaluation*. ([www.alnap.org/resource/22912.aspx](http://www.alnap.org/resource/22912.aspx)).

**Davidson, E. J. (2012a)** '6 questions that cut to the chase when choosing the right evaluation team', in Rogers, P. J. and Davidson, E. J. (eds) *Genuine Evaluation*. ([www.alnap.org/resource/22905.aspx](http://www.alnap.org/resource/22905.aspx)).

**Davidson, E. J. (2012b)** '9 hot tips for commissioning, managing (and doing!) actionable evaluation' in Rogers, P. J. and Davidson, E. J. (eds) *Genuine Evaluation*. ([www.alnap.org/resource/22913.aspx](http://www.alnap.org/resource/22913.aspx)).

**Davidson, E. J. (2014a)** *Evaluative Reasoning*. Florence: UNICEF. ([www.alnap.org/resource/22331.aspx](http://www.alnap.org/resource/22331.aspx)).

**Davidson, E. J. (2014b)** "'Minirubrics' – 7 hot tips for using this cool tool to focus evaluative conversations', in Rogers, P. J. and Davidson, E. J. (eds) *Genuine Evaluation*. ([www.alnap.org/resource/22914.aspx](http://www.alnap.org/resource/22914.aspx)).

**Davies, R. and Dart, J. (2005)** *The "most significant change" (MSC) technique: A guide to its use*. Atlanta, GA: CARE International. ([www.alnap.org/resource/8102.aspx](http://www.alnap.org/resource/8102.aspx)).

**Davin, E., Gonzalez, V. and Majidi, N. (2009)** *UNHCR's voluntary repatriation program: Evaluation of the impact of the cash grant*. Geneva: UNHCR. ([www.alnap.org/resource/22915.aspx](http://www.alnap.org/resource/22915.aspx)).

**Davis, R. E., Couper, M. P., Janz, N. K., Caldwell, C. H. and Resnicow, K. (2010)** 'Interviewer effects in public health surveys'. *Health Education Research*, 25(1): 14–26. ([www.alnap.org/resource/8103.aspx](http://www.alnap.org/resource/8103.aspx)).

**Day, S. J. and Altman, D. G. (2000)** 'Blinding in clinical trials and other studies'. *BMJ*, 321(8): 504. ([www.alnap.org/resource/22917.aspx](http://www.alnap.org/resource/22917.aspx)).

**DEC. (2013)** *DEC Approach to Learning*. London: DEC. ([www.alnap.org/resource/8649.aspx](http://www.alnap.org/resource/8649.aspx)).

**de Nicola, F. and Giné, X. (2011)** *How accurate are recall data? Evidence from coastal India*. Washington, DC: World Bank. ([www.alnap.org/resource/8104.aspx](http://www.alnap.org/resource/8104.aspx)).

**Denzin, N. K. (2010)** 'Moments, mixed methods, and paradigm dialogs'. *Qualitative Inquiry*, 16(6): 419–427. ([www.alnap.org/resource/8105.aspx](http://www.alnap.org/resource/8105.aspx)).

**de Ville de Goyet, C. and Morinière, L. (2006)** *The role of needs assessment in the tsunami response*. London: TEC. ([www.alnap.org/resource/12406.aspx](http://www.alnap.org/resource/12406.aspx)).

**de Ville de Goyet, C., Scheuermann, P., Reed, S. B., Al Wehaidy, R., and Termil, A. (2011)** *SDC humanitarian aid: Emergency relief*. Bern: Swiss Agency for Development and Cooperation. ([www.alnap.org/resource/6142.aspx](http://www.alnap.org/resource/6142.aspx)).

**de Ville de Goyet, C., Buckle, P., Ievers, J., Parlan, H. P. and Viandrito, J. (2013)** *Evaluation of the European Commission's humanitarian and disaster risk reduction activities (DIPECHO) in Indonesia*. Brussels: ECHO. ([www.alnap.org/resource/11739.aspx](http://www.alnap.org/resource/11739.aspx)).

**DFID. (2003)** *Tools for development: A handbook for those engaged in development activity*. London: DFID. ([www.alnap.org/resource/11859.aspx](http://www.alnap.org/resource/11859.aspx)).

**DFID, (2005)** *Guidance and evaluation and review for DFID staff*. London: DFID. ([www.alnap.org/resource/22918.aspx](http://www.alnap.org/resource/22918.aspx)).

**DFID. (2010)** *Evaluation study terms of reference (template)*. London: DFID. ([www.alnap.org/resource/22920.aspx](http://www.alnap.org/resource/22920.aspx)).

**DFID. (2014)** *Assessing the strength of evidence*. London: DFID. ([www.alnap.org/resource/10949.aspx](http://www.alnap.org/resource/10949.aspx)).

**DFID and UKAID. (2011)** *Multilateral aid review: ensuring maximum value for money for UK aid through multilateral organisations*. London: DFID and UKAID. ([www.alnap.org/resource/8106.aspx](http://www.alnap.org/resource/8106.aspx)).

**Dillman, D. A., Smyth, J. D. and Christian, L. M. (2009)** *Internet, mail and mixed-mode surveys: The tailored design method*. 3rd ed. Hoboken, NJ: Wiley. ([www.alnap.org/resource/8107.aspx](http://www.alnap.org/resource/8107.aspx)).

**Disch, A., Haarberg, K., Gharthey, A. B. and Lunøe, B. (2008)** *Synthesis study on best practices and innovative approaches to capacity development in low-income African countries*. Oslo: NORAD. ([www.alnap.org/resource/10132.aspx](http://www.alnap.org/resource/10132.aspx)).

**Dohrenwend, B. S., Colombotos, J. and Dohrenwend, B. P. (1968)** 'Social distance and interviewer effects'. *Public Opinion Quarterly*, 32(3): 410–422. ([www.alnap.org/resource/8109.aspx](http://www.alnap.org/resource/8109.aspx)).

**Drucker, P. F. (1985)** *Innovation and entrepreneurship*. Oxford: Butterworth-Heinemann. ([www.alnap.org/resource/22921.aspx](http://www.alnap.org/resource/22921.aspx)).

**Dryden-Peterson, S. (2011)** *Refugee education: A global review*. Geneva: UNHCR. ([www.alnap.org/resource/8110.aspx](http://www.alnap.org/resource/8110.aspx)).

**Duncalf, J. (2013)** *A real-time evaluation of ACF International's response to typhoon Haiyan / Yolanda in the Philippines*. Paris: ACF. ([www.alnap.org/resource/20994.aspx](http://www.alnap.org/resource/20994.aspx)).

**du Preez, M., Conroy, R. M., Wright, J. A., Moyo, S., Potgieter, N. and Gundry, S. W. (2008)** 'Use of ceramic water filtration in the prevention of diarrheal disease: A randomized controlled trial in rural South Africa and Zimbabwe. *The American Journal of Tropical Medicine and Hygiene*, 79(5): 696–701. ([www.alnap.org/resource/22922.aspx](http://www.alnap.org/resource/22922.aspx)).

**Durkin, M. S., Khan, N., Davidson, L. L., Zaman, S. S. and Stein, Z. A. (1993)** 'The effects of a natural disaster on child behavior: evidence for posttraumatic stress'. *American Journal of Public Health*, 83(11): 1549–1553. ([www.alnap.org/resource/12530.aspx](http://www.alnap.org/resource/12530.aspx)).

**Ebbinghaus, B. (2005)** 'When less is more: Selection problems in large- N and small- N cross-national comparisons. *International Sociology*, 20(2): 133–152. ([www.alnap.org/resource/22924.aspx](http://www.alnap.org/resource/22924.aspx)).

**ECHO. (2002)** *Evaluating humanitarian action funded by the humanitarian aid office of the European Commission*. Brussels: ECHO. ([www.alnap.org/resource/22925.aspx](http://www.alnap.org/resource/22925.aspx)).

**Edmondson, A. C. (2004)** 'Learning from failure in health care: frequent opportunities, pervasive barriers'. *Quality and Safety in Health Care*, 13(2):ii3–ii9. ([www.alnap.org/resource/8111.aspx](http://www.alnap.org/resource/8111.aspx)).

**ECB. (2011a)** *What we know about joint evaluations of humanitarian action*. London: ECB. ([www.alnap.org/resource/19286.aspx](http://www.alnap.org/resource/19286.aspx)).

**ECB. (2011b)** *What we know about joint evaluations of humanitarian action. Section 1 of 3: the guide*. London: ECB. ([www.alnap.org/resource/9536.aspx](http://www.alnap.org/resource/9536.aspx)).

**ECB. (2011c)** *What we know about joint evaluations of humanitarian action. Section 2 of 3: the stories*. London: ECB. ([www.alnap.org/resource/9537.aspx](http://www.alnap.org/resource/9537.aspx)).

**ECB. (2011d)** *What we know about joint evaluations of humanitarian action. Section 3 of 3: the tools*. London: ECB. ([www.alnap.org/resource/9538.aspx](http://www.alnap.org/resource/9538.aspx)).

**ECB. (2012)** *Joint evaluations. Guidelines from the ECB project*. London: ECB. ([www.alnap.org/resource/19413.aspx](http://www.alnap.org/resource/19413.aspx)).

**ECB. (2013)** *What we know about collaboration: the ECB country consortium experience*. London: ECB. ([www.alnap.org/resource/10552.aspx](http://www.alnap.org/resource/10552.aspx)).

**EES. (2007)** *EES statement: The importance of a methodologically diverse approach to impact evaluation - specifically with respect to development aid and development interventions*. Prague: EES. ([www.alnap.org/resource/11119.aspx](http://www.alnap.org/resource/11119.aspx)).

**EES. (2015)** *The EES capabilities framework*. Prague: EES. ([www.alnap.org/resource/22927.aspx](http://www.alnap.org/resource/22927.aspx)).

**EPOC. (2013)** *Interrupted time series (ITS) analyses*. Oslo: Norwegian Knowledge Centre for the Health Services. ([www.alnap.org/resource/22928.aspx](http://www.alnap.org/resource/22928.aspx)).

**Egeland, J., Harmer, A. and Stoddard, A. (2011)** *To stay and deliver: Good practice for humanitarian in complex security environments*. New York: OCHA. ([www.alnap.org/resource/6364.aspx](http://www.alnap.org/resource/6364.aspx)).

**Eisinga, R., te Grotenhuis, M., Larsen, J. K., Pelzer, B. and van Strien, T. (2011)** BMI of interviewer effects. *International Journal of Public Opinion Research*, 23(4): 530–543. ([www.alnap.org/resource/8112.aspx](http://www.alnap.org/resource/8112.aspx)).

**FDFA. (2014)** *Humanitarian access in situations of armed conflict*. Bern: FDFA. ([www.alnap.org/resource/21733.aspx](http://www.alnap.org/resource/21733.aspx)).

**Fearon, J., Humphreys, M. and Weinstein, J. (2008)** *Community driven reconstruction in Lofa County: Impact assessment*. New York: IRC. ([www.alnap.org/resource/8192.aspx](http://www.alnap.org/resource/8192.aspx)).

**Featherstone, A. (2012)** *Evaluation of Somalia drought response 2011/2012: Using Oxfam GB's global humanitarian indicator tool*. Oxford: Oxfam. ([www.alnap.org/resource/7983.aspx](http://www.alnap.org/resource/7983.aspx)).

**Featherstone, A. (2013a)** *Improving impact: do accountability mechanisms deliver results?* Geneva: HAP. ([www.alnap.org/resource/8388.aspx](http://www.alnap.org/resource/8388.aspx)).

**Featherstone, A. (2013b)** *Syria needs analysis project (SNAP): External mid-term review*. Geneva: ACAPS. ([www.alnap.org/resource/10390.aspx](http://www.alnap.org/resource/10390.aspx)).

**Featherstone, A. (2014)** *A review of UNHCR's utilisation of the Central Emergency Response Fund*. Geneva: UNHCR. ([www.alnap.org/resource/12430.aspx](http://www.alnap.org/resource/12430.aspx)).

**Ferf, A. and Fabbri, P. (2014)** *Impact evaluation of Swiss solidarity Asian tsunami programme*. Geneva: Swiss Solidarity. ([www.alnap.org/resource/22473.aspx](http://www.alnap.org/resource/22473.aspx)).

- Fergusson, D., Cranley Glass, K., Waring, D., Shapiro, S. (2004)** 'Turning a blind eye: the success of blinding reported in a random sample of randomised, placebo controlled trials'. *BMJ*. ([www.alnap.org/resource/22930.aspx](http://www.alnap.org/resource/22930.aspx)).
- Few, R., McAvoy, D., Tarazona, M. and Walden, V. (2014)** *Contribution to Change: An approach to evaluating the role of intervention in disaster recovery*. Oxford: Oxfam. ([www.alnap.org/resource/10305.aspx](http://www.alnap.org/resource/10305.aspx)).
- Flaherty, E. W., Barry, E. and Swift, M. (1978)** 'Use of an unobtrusive measure for the evaluation of interagency coordination. *Evaluation Review*, 2(2): 261–273. ([www.alnap.org/resource/22931.aspx](http://www.alnap.org/resource/22931.aspx)).
- Flores-Macias, F. and Lawson, C. (2008)** 'Effects of interviewer gender on survey responses: Findings from a household survey in Mexico'. *International Journal of Public Opinion Research*, 20(1): 100–110. ([www.alnap.org/resource/8113.aspx](http://www.alnap.org/resource/8113.aspx)).
- Foley, P. (2009)** *Participatory evaluation of the 2008 farmer field school programme, Lira, Uganda*. Kampala: ACF. ([www.alnap.org/resource/20908.aspx](http://www.alnap.org/resource/20908.aspx)).
- Fortune, V. and Rasal, P. (2010)** *British Red Cross - mass sanitation module - 2010 Haiti earthquake response - post deployment learning evaluation*. London: British Red Cross. ([www.alnap.org/resource/6038.aspx](http://www.alnap.org/resource/6038.aspx)).
- Foubert, V. and Eskandar, H. (2009)** *Taking the initiative – Exploring quality and accountability in the humanitarian sector: an introduction to eight initiatives*. Geneva: Sphere Project. ([www.alnap.org/resource/5673.aspx](http://www.alnap.org/resource/5673.aspx)).
- Fowler, F. J. (2009)** *Survey research methods*. Thousand Oaks, CA: SAGE. ([www.alnap.org/resource/8115.aspx](http://www.alnap.org/resource/8115.aspx)).
- Fritz Institute. (2005)** *Recipient perceptions of aid effectiveness: rescue, relief and rehabilitation in tsunami affected Indonesia, India and Sri Lanka*. San Francisco: CA: Fritz Institute. ([www.alnap.org/resource/5459.aspx](http://www.alnap.org/resource/5459.aspx)).
- Fritz Institute. (2006)** *Hurricane Katrina: Perceptions of the affected*. San Francisco, CA: Fritz Institute. ([www.alnap.org/resource/8116.aspx](http://www.alnap.org/resource/8116.aspx)).
- Frongillo, E. A. (2012)** *Analysis of complex surveys*. Cornell: Cornell Statistical Consulting Unit. ([www.alnap.org/resource/22932.aspx](http://www.alnap.org/resource/22932.aspx)).

- Funnell, S. C. and Rogers, P. (2011)** *Purposeful program theory: Effective use of theories of change and logic models*. Hoboken, NJ: Wiley. ([www.alnap.org/resource/8193.aspx](http://www.alnap.org/resource/8193.aspx)).
- Gamarra, T., Reed, S. B. and Wilding, J. (2005)** *Final evaluation of hurricanes operation 2004*. Geneva: IFRC. ([www.alnap.org/resource/12536.aspx](http://www.alnap.org/resource/12536.aspx)).
- Garvin, D. A., Edmondson, A. C. and Gino, F. (2008)** 'Is yours a learning organization?' *Harvard Business Review*, 86(3). ([www.alnap.org/resource/8117.aspx](http://www.alnap.org/resource/8117.aspx)).
- Gerring, J. (2006)** 'Techniques for choosing cases', in *Case study research: principles and practices*. New York: Cambridge University Press. ([www.alnap.org/resource/9014.aspx](http://www.alnap.org/resource/9014.aspx)).
- Gilmour, S., Degenhardt, L., Hall, W. and Day, C. (2006)** 'Using intervention time series analyses to assess the effects of imperfectly identifiable natural events: a general method and example'. *BMC Medical Research Methodology*, 6(1): 1–9. ([www.alnap.org/resource/22933.aspx](http://www.alnap.org/resource/22933.aspx)).
- Given, L. M. (2008)** *The SAGE encyclopedia of qualitative research methods*. Thousand Oaks, CA: SAGE. ([www.alnap.org/resource/22934.aspx](http://www.alnap.org/resource/22934.aspx)).
- Glass, G. V. (1997)** *Interrupted time series quasi-experiments*. Tempe, AZ: Arizona State University. ([www.alnap.org/resource/22935.aspx](http://www.alnap.org/resource/22935.aspx)).
- Glouberman, S. and Zimmerman, B. (2002)** *Complicated and complex systems: what would successful reform of Medicare look like?* Ottawa: Commission on the Future of Health Care in Canada. ([www.alnap.org/resource/8119.aspx](http://www.alnap.org/resource/8119.aspx)).
- Godden, B. (2004)** *Sample size formulas*. Chicago: Appraisal institute. ([www.alnap.org/resource/22936.aspx](http://www.alnap.org/resource/22936.aspx)).
- Goertz, G. (2012)** *A tale of two cultures: Qualitative and quantitative research in the social sciences*. Princeton: Princeton University Press. ([www.alnap.org/resource/13029.aspx](http://www.alnap.org/resource/13029.aspx)).
- Goertz, G. and Mahoney, J. (2012)** 'Methodological Rorschach tests: Contrasting interpretations in qualitative and quantitative research'. *Comparative Political Studies*, 46(2): 236–251. ([www.alnap.org/resource/22937.aspx](http://www.alnap.org/resource/22937.aspx)).

**GHD. (2003a)** 'Good practices in donor financing, management and accountability', in *23 principles and good practice of humanitarian donorship*. (www.alnap.org/resource/10184.aspx).

**GHD. (2003b)** *International meeting on good humanitarian donorship*. Stockholm: GHD. (www.alnap.org/resource/22940.aspx).

**Goodrick, D. (2014)** *Comparative case studies*. Florence: UNICEF. (www.alnap.org/resource/22941.aspx).

**Gover, A. R., Macdonald, J. M. and Alpert, G. P. (2003)** 'Combating domestic violence: Findings from an evaluation of a local domestic violence court'. *Criminology & Public Policy*, 3(1): 109–132. (www.alnap.org/resource/22942.aspx).

**Goyder, H. (2010)** *Evaluation of DEC-funded shelter projects following the 2009 Indonesian earthquake*. Teddington: Tearfund. (www.alnap.org/resource/5958.aspx).

**Graig, E. (2011)** *Hiring an evaluation consultant*. Riverdale: Usable Knowledge. (www.alnap.org/resource/22944.aspx).

**Graig, E. (2013)** 'An alternative to the evaluation RFP process', in Usable Knowledge. (www.alnap.org/resource/22943.aspx).

**Grais, R. F., Luquero, F. J., Grellety, E., Pham, H., Coghlan, B. and Salignon, P. (2009)** 'Learning lessons from field surveys in humanitarian contexts: a case study of field surveys conducted in North Kivu, DRC 2006-2008'. *Conflict and Health*, 3(8). (www.alnap.org/resource/22946.aspx).

**Grasso, P. G. Imas, L. M. and Fostvedt, N. (2013)** *Use of evaluation in the Norwegian development cooperation system*. Oslo: Norad. (www.alnap.org/resource/22947.aspx).

**Gray, D.E. (2014)** *Doing research in the real world*. 3rd ed. Los Angeles: SAGE. (www.alnap.org/resource/22948.aspx).

**Green, D. (2012)** *Creating killer facts and graphics*. Oxford: Oxfam. (www.alnap.org/resource/22949.aspx).

**Groupe URD. (2009)** *Quality Compas Companion Book*. Plaisians: Groupe URD. (www.alnap.org/resource/8541.aspx).

- Grünewald, F., Binder, A. and Georges, Y. (2010)** *Inter-agency real-time evaluation in Haiti: 3 months after the earthquake*. Plaisians: Groupe URD. ([www.alnap.org/resource/9962.aspx](http://www.alnap.org/resource/9962.aspx)).
- Grünewald, F., Kauffmann, D., Boyer, B. and Patinet, J. (2011)** *Real-time evaluation of humanitarian action supported by DG ECHO in Haiti: 2009-2011: November 2010-April 2011*. Plaisians: Groupe URD. ([www.alnap.org/resource/22950.aspx](http://www.alnap.org/resource/22950.aspx)).
- Gubbels, P. and Bousquet, C. (2013)** *Independent evaluation of CARE's response to the 2011-2012 Sahel humanitarian crisis*. Geneva: CARE International. ([www.alnap.org/resource/20768.aspx](http://www.alnap.org/resource/20768.aspx)).
- Guerrero, S., Woodhead, S. and Hounjet, M. (2013)** *On the right track: a brief review of monitoring and evaluation in the humanitarian sector*. ([www.alnap.org/resource/8211.aspx](http://www.alnap.org/resource/8211.aspx)).
- Guest, G., Bunce, A. and Johnson, L. (2006)** How many interviews are enough? *Field methods*, 18(1): 59–82. ([www.alnap.org/resource/22952.aspx](http://www.alnap.org/resource/22952.aspx)).
- Guijt, I. (2014)** *Participatory Approaches*, Florence: UNICEF. ([www.alnap.org/resource/22798.aspx](http://www.alnap.org/resource/22798.aspx)).
- Hagens, C. and Ishida, L. (2010)** *Real time evaluation of CRS's flood response in Pakistan: KPK and Baluchistan*. Baltimore, MD: CRS. ([www.alnap.org/resource/5956.aspx](http://www.alnap.org/resource/5956.aspx)).
- Hahn, J., Todd, P. and Van der Klaauw, W. (2001)** 'Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1): 201–209. ([www.alnap.org/resource/22954.aspx](http://www.alnap.org/resource/22954.aspx)).
- Hallam, A. (2011)** *Harnessing the power of evaluation in humanitarian action: An initiative to improve understanding and use of evaluation*. London: ALNAP. ([www.alnap.org/resource/6123.aspx](http://www.alnap.org/resource/6123.aspx)).
- Hallam, A. and Bonino, F. (2013)** *Using evaluation for a change: Insights from humanitarian practitioners*. London: ALNAP. ([www.alnap.org/resource/8980.aspx](http://www.alnap.org/resource/8980.aspx)).
- Hammersley, M. (1992)** *What's wrong with ethnography?* London: Routledge. ([www.alnap.org/resource/9017.aspx](http://www.alnap.org/resource/9017.aspx)).

- Hanley, T., Binas, R., Murray, J. and Tribunalo, B. (2014)** *IASC Inter-agency humanitarian evaluation of the typhoon Haiyan response*. Geneva: IASC. ([www.alnap.org/resource/19318.aspx](http://www.alnap.org/resource/19318.aspx)).
- HAP. (2010)** *The 2010 HAP standard in accountability and quality management*. Geneva: HAP. ([www.alnap.org/resource/8125.aspx](http://www.alnap.org/resource/8125.aspx)).
- Hart, C. (1998)** *Doing a literature review: releasing the social science research imagination*. Thousand Oaks, CA: SAGE. ([www.alnap.org/resource/8126.aspx](http://www.alnap.org/resource/8126.aspx)).
- Hassan, P., Freese, A. and Guallar, M. (2013)** *Desk review of UNDAFs commencing in 2013*. New York: UNDP. ([www.alnap.org/resource/22811.aspx](http://www.alnap.org/resource/22811.aspx)).
- Haver, K., Hatungimana, F. and Tennant, V. (2009)** *Money matters: An evaluation of the use of cash grants in UNHCR's voluntary repatriation and reintegration programme in Burundi*. Geneva: UNHCR. ([www.alnap.org/resource/5710.aspx](http://www.alnap.org/resource/5710.aspx)).
- HelpAge International. (2010)** *A study of humanitarian funding for older people*. London: HelpAge International. ([www.alnap.org/resource/8127.aspx](http://www.alnap.org/resource/8127.aspx)).
- Hidalgo, S., LaGuardia, D., Trudi, G., Sole, R., Moussa, Z., van Dijk, J., Merckx, P. and Zimmer, L. (2015)** *Beyond humanitarian assistance? UNHCR and the response to Syrian refugees in Jordan and Lebanon: January 2013 – April 2014*. Brussels: Transtec. ([www.alnap.org/resource/20222.aspx](http://www.alnap.org/resource/20222.aspx)).
- Hidalgo, S. and Théodate, M. P. (2011)** *Inter-agency real-time evaluation of the humanitarian response to the earthquake in Haiti: 20 months after*. Geneva: IASC. ([www.alnap.org/resource/6330.aspx](http://www.alnap.org/resource/6330.aspx)).
- Hill, M. E. (2002)** 'Race of the interviewer and perception of skin color: Evidence from the multi-city study of urban inequality'. *American Sociological Review*, 67(1): 99–108. ([www.alnap.org/resource/8123.aspx](http://www.alnap.org/resource/8123.aspx)).
- Horton, K. and Roche, C. (eds) (2010)** *Ethical questions and international NGOs: An exchange between philosophers and NGOs*. Houten: Springer. ([www.alnap.org/resource/22955.aspx](http://www.alnap.org/resource/22955.aspx)).
- Hróbjartsson, A., Forfang, E., Haahr, M. T., Als-Nielsen, B. and Brorson, S. (2007)** 'Blinded trials taken to the test: an analysis of randomized clinical trials that report tests for the success of blinding'. *International Journal of Epidemiology*, 36(3): 654–663. ([www.alnap.org/resource/22956.aspx](http://www.alnap.org/resource/22956.aspx)).

**Humanitarian Initiatives UK. (2001)** *Independent evaluation: The DEC response to the earthquake in Gujarat January - October 2001*. London: Humanitarian Initiatives UK. ([www.alnap.org/resource/3432.aspx](http://www.alnap.org/resource/3432.aspx)).

**Hund, L., Bedrick, E. J. and Pagano, M. (2015)** 'Choosing a cluster sampling design for lot quality assurance sampling surveys. *PLoS ONE*, 10(6). ([www.alnap.org/resource/22958.aspx](http://www.alnap.org/resource/22958.aspx)).

**Hunt, D. (2014)** *Photo Story: Risk Analysis in Tcharow, Goz Beida, Sila Region, Chad*. Concern Worldwide.

**Hyatt, S. and Auten, S. (2011)** *Evaluation of the Palestinian youth empowerment program (RUWWAD)*. Washington, DC: USAID. ([www.alnap.org/resource/22959.aspx](http://www.alnap.org/resource/22959.aspx)).

**IASC. (2010)** *Gender marker: How-to-code tip sheet*. Geneva: IASC. ([www.alnap.org/resource/8128.aspx](http://www.alnap.org/resource/8128.aspx)).

**IASC. (2010b)** *Guidance note for clusters to implement the IASC gender marker: Creating gender-responsive projects and tracking gender-related allocations in humanitarian appeals and funding mechanisms*. Geneva: IASC. ([www.alnap.org/resource/8129](http://www.alnap.org/resource/8129)).

**IASC. (2010c)** *IASC gender marker fact sheet*. Geneva: IASC. ([www.alnap.org/resource/22800.aspx](http://www.alnap.org/resource/22800.aspx)).

**IASC. (2010d)** *Inter-agency real-time evaluation (IA RTE) of the humanitarian response to [disaster XYZ In country XYZ]: Terms of reference*. Geneva: IASC. ([www.alnap.org/resource/22960.aspx](http://www.alnap.org/resource/22960.aspx)).

**IASC. (2011)** *IASC operational guidelines on protection of persons in situations of natural disasters*. Geneva: IASC. ([www.alnap.org/resource/12828.aspx](http://www.alnap.org/resource/12828.aspx)).

**IASC. (2013)** *The centrality of protection in humanitarian action*. Geneva: IASC. ([www.alnap.org/resource/19093.aspx](http://www.alnap.org/resource/19093.aspx)).

**ICAI. (2013)** *DFID's support for Palestine refugees through UNRWA*. London: ICAI. ([www.alnap.org/resource/10111.aspx](http://www.alnap.org/resource/10111.aspx)).

**ICRC. (1977)** *Protocol additional to the Geneva conventions of 12 August 1949, and relating to the protection of victims of non-international armed conflicts (Protocol II)*. Geneva: ICRC. ([www.alnap.org/resource/8122.aspx](http://www.alnap.org/resource/8122.aspx)).

**IFAD. (2011)** *Evaluating the impact of participatory mapping activities*. Rome: IFAD. ([www.alnap.org/resource/22962.aspx](http://www.alnap.org/resource/22962.aspx)).

**IFAD. (2014)** *Calculating the sample size*. Rome: IFAD. ([www.alnap.org/resource/22961.aspx](http://www.alnap.org/resource/22961.aspx)).

**IFRC (n.d.)** *Types of disasters*. Geneva: IFRC. ([www.alnap.org/resource/22963.aspx](http://www.alnap.org/resource/22963.aspx)).

**IFRC. (2011)** *IFRC framework for evaluation*. Geneva: IFRC. ([www.alnap.org/resource/7975.aspx](http://www.alnap.org/resource/7975.aspx)).

**IFRC. (2010)** *World disasters report 2010: Focus on urban risk*. Geneva: IFRC. ([www.alnap.org/resource/9272.aspx](http://www.alnap.org/resource/9272.aspx)).

**IFRC. (2011)** *Project/programme monitoring and evaluation (M&E) guide*. Geneva: IFRC. ([www.alnap.org/resource/8542.aspx](http://www.alnap.org/resource/8542.aspx)).

**IFRC and ICRC. (1994)** *Code of conduct for the International Red Cross and Red Crescent movement and non-governmental organizations (NGOs) in disaster relief*. Geneva: IFRC/ICRC. ([www.alnap.org/resource/19291.aspx](http://www.alnap.org/resource/19291.aspx)).

**3ie. (2008a)** *3ie impact evaluation practice: a guide for grantees*. New Delhi: 3ie ([www.alnap.org/resource/8130.aspx](http://www.alnap.org/resource/8130.aspx)).

**3ie. (2008b)** *Principles for impact evaluation*. New Delhi: 3ie. ([www.alnap.org/resource/8131.aspx](http://www.alnap.org/resource/8131.aspx)).

**ILO. (2014a)** *Checklist 5 preparing the evaluation report*. Geneva: ILO. ([www.alnap.org/resource/22964.aspx](http://www.alnap.org/resource/22964.aspx)).

**ILO. (2014b)** *Checklist 6 rating the quality of evaluation*. Geneva: ILO. ([www.alnap.org/resource/22965.aspx](http://www.alnap.org/resource/22965.aspx)).

**Imbens, G. W. and Lemieux, T. (2007)** 'Regression discontinuity designs: A guide to practice'. *Journal of Econometrics*, 142(2): 615–635. ([www.alnap.org/resource/22966.aspx](http://www.alnap.org/resource/22966.aspx)).

**INEE. (2004)** *Minimum standards for education in emergencies, chronic crises and early reconstruction*. New York: INEE. ([www.alnap.org/resource/7860.aspx](http://www.alnap.org/resource/7860.aspx)).

**Innovation Network. (2005)** *Logic model workbook*. Washington, DC: Innovation Network. ([www.alnap.org/resource/22967.aspx](http://www.alnap.org/resource/22967.aspx)).

**Israel, G. D. (2009)** *Determining sample size*. Gainesville: IFAS University of Florida. ([www.alnap.org/resource/22968.aspx](http://www.alnap.org/resource/22968.aspx)).

**Jadad, A. R., Moore, A., Carroll, D., Jenkinson, C., Reynolds, J. M., Gavaghan, D. J. and McQuay, H. J. (1996)** 'Assessing the quality of reports of randomized clinical trials: is blinding necessary?' *Controlled Clinical Trials*, 17(1): 1–12. ([www.alnap.org/resource/22969.aspx](http://www.alnap.org/resource/22969.aspx)).

**Jalan, J. and Ravallion, M. (2001)** 'Does piped water reduce diarrhea for children in rural India?' *Journal of Econometrics*, 112(1): 153-173. ([www.alnap.org/resource/22970.aspx](http://www.alnap.org/resource/22970.aspx)).

**Jaspars, S. and Young, H. (1995)** *General food distribution in emergencies*, London: ODI. ([www.alnap.org/resource/2725.aspx](http://www.alnap.org/resource/2725.aspx)).

**Johnson, A. and Head, D. (2011)** *CRS Haiti real time evaluation of the 2010 earthquake response : findings, recommendations, and suggested follow up*. Baltimore, MD: CRS. ([www.alnap.org/resource/6074.aspx](http://www.alnap.org/resource/6074.aspx)).

**Johri, M., Ridde, V., Heinmüller, R. and Haddad, S. (2014)** *Estimation of maternal and child mortality one year after user-fee elimination: an impact evaluation and modelling study in Burkina Faso*. Geneva: WHO. ([www.alnap.org/resource/22971.aspx](http://www.alnap.org/resource/22971.aspx)).

**Jones, N., Jones, H., Steer, L. and Datta, A. (2009)** *Improving impact evaluation production and use*. London: ODI. ([www.alnap.org/resource/8225.aspx](http://www.alnap.org/resource/8225.aspx)).

**Journal of Peacebuilding & Development. (2013)** 'Evaluation in violently divided societies: Politics, ethics and methods'. *Journal of Peacebuilding & Development*, 8(2): 1–4. ([www.alnap.org/resource/23078.aspx](http://www.alnap.org/resource/23078.aspx)).

**Kahneman, D., Knetsch, J. L. and Thaler, R. H. (1991)** 'The endowment effect, loss aversion, and status quo bias'. *The Journal of Economic Perspectives*, 5(1): 193–206. ([www.alnap.org/resource/22972.aspx](http://www.alnap.org/resource/22972.aspx)).

**Kaiser, R., Woodruff, B. A., Bilukha, O., Spiegel, P. B. and Salama, P. (2006)** 'Using design effects from previous cluster surveys to guide sample size calculation in emergency settings'. *Disasters*, 30(2): 199–211. ([www.alnap.org/resource/22973.aspx](http://www.alnap.org/resource/22973.aspx)).

**Kalantari, M., Yule, W., Dyregrov, A., Neshatdoost, H. and Ahmadi, S. J. (2012)** 'Efficacy of writing for recovery on traumatic grief symptoms of Afghani refugee bereaved adolescents: A randomized control trial. *OMEGA - Journal of Death and Dying*, 65(2): 139–150. ([www.alnap.org/resource/22974.aspx](http://www.alnap.org/resource/22974.aspx)).

**Kane, E. W. and Macaulay, L. J. (1993)** 'Interviewer gender and gender attitudes'. *Public Opinion Quarterly*, 57(1): 1–28. ([www.alnap.org/resource/8132.aspx](http://www.alnap.org/resource/8132.aspx)).

**Kaplan, A. M. and Haenlein, M. (2010)** 'Users of the world, unite! The challenges and opportunities of social media'. *Business Horizons*, 53(1): 59–68. ([www.alnap.org/resource/22975.aspx](http://www.alnap.org/resource/22975.aspx)).

**Kennedy, M. (2006)** *Interview Probes*. ([www.alnap.org/resource/22976.aspx](http://www.alnap.org/resource/22976.aspx)).

**Kevlihan, R. (2010)** Productivity and cash-for-work in Niger: GOAL's experience. *Humanitarian Exchange* 48: 29-30. London: HPN/ODI. ([www.alnap.org/resource/8135.aspx](http://www.alnap.org/resource/8135.aspx)).

**Khandker, S. R., Koolwal, G. B. and Samad, H. A. (2010)** *Handbook on impact evaluation: quantitative methods and practices*. Washington, DC: World Bank. ([www.alnap.org/resource/22977.aspx](http://www.alnap.org/resource/22977.aspx)).

**Khoo, S.-E. (2010)** 'Health and humanitarian migrants' economic participation. *Journal of Immigrant and Minority Health*, 12(3): 327–339. ([www.alnap.org/resource/22978.aspx](http://www.alnap.org/resource/22978.aspx)).

**King, J., McKegg, K., Oakden, J. and Wehipeihana, N. (2013)** 'Rubrics: A method for surfacing values and improving the credibility of evaluation'. *Journal of multidisciplinary evaluation*, 9(21): 11–20. ([www.alnap.org/resource/22424.aspx](http://www.alnap.org/resource/22424.aspx)).

**Kirkby, J., Saeed, A. and Zogopoulos, A. (2006)** *Independent evaluation of CARE international's earthquake response in Northern Pakistan*. Atlanta, GA: CARE International. ([www.alnap.org/resource/3489.aspx](http://www.alnap.org/resource/3489.aspx)).

**Knox Clarke, P. and Darcy, J. (2014)** *Insufficient evidence? The quality and use of evidence in humanitarian action*. ALNAP Study. London: ALNAP/ODI. ([www.alnap.org/resource/10441.aspx](http://www.alnap.org/resource/10441.aspx)).

**Koenig, T. (2004)** *CAQDAS - A primer*. ([www.alnap.org/resource/22979.aspx](http://www.alnap.org/resource/22979.aspx)).

**Kohrt, B. A., Burkey, M., Stuart, E., A. and Koirala, S. (2015)** 'Alternative approaches for studying humanitarian interventions: propensity score methods to evaluate reintegration packages impact on depression, PTSD, and function impairment among child soldiers in Nepal'. *Global Mental Health*. (www.alnap.org/resource/22980.aspx).

**Kosko, J., Klassen, T. P., Bishop, T. and Hartling, L. (2006)** 'Evidence-based medicine and the anecdote: Uneasy bedfellows or ideal couple?' *Paediatrics & Child Health*, 11(10): 665–668. (www.alnap.org/resource/22981.aspx).

**Krippendorff, K. (2004)** *Content analysis: an introduction to its methodology*. 2nd ed. Thousand Oaks, CA: SAGE. (www.alnap.org/resource/8223.aspx).

**Krueger, R. A. and Casey, M. A. (2009)** *Focus groups: A practical guide for applied research*. Thousand Oaks, CA: Sage. (www.alnap.org/resource/8136.aspx).

**Krueger, S. and Sagmeister, E. (2014)** *Real-time evaluation of humanitarian assistance revisited: Lessons learned and the way forward*. Berlin: GPPi. (www.alnap.org/resource/12886.aspx).

**Kumar, S. (2002)** *Methods for community participation: a complete guide for practitioners*. Rugby: ITDG Publishing. (www.alnap.org/resource/8137.aspx).

**Kusters, C., van Vugt, S., Wigboldus, S., Williams, B. and Woodhill, J. (2011)** *Making evaluations matter: A practical guide for evaluators*. (www.alnap.org/resource/22425.aspx).

**Ky Luu, J. D., Nkwake, A. M., Morris, M. F., Hansch, S., Gago, M., Corzine, E. and Marcello, J. (2014)** *A decade of learning: Lessons from an evaluation of the emergency capacity building (ECB) project*. New Orleans, LA: DRG. (www.alnap.org/resource/12981.aspx).

**LaGuardia, D. and Van den Toorn, W. (2011)** *Evaluation of UNRWA's organizational development (OD)*. Amman: UNRWA. (www.alnap.org/resource/22986.aspx).

**Lanphear, B. P., Vorhees, C. V. and Bellinger, D. C. (2005)** 'Protecting children from environmental toxins'. *PLoS Medicine*, 2(3): 203-208. (www.alnap.org/resource/22987.aspx).

- Larcom, S., Rauch, F. and Willems, T. (2015)** *The benefits of forced experimentation: Striking evidence from the London Underground network*, Oxford: University of Oxford. ([www.alnap.org/resource/22988.aspx](http://www.alnap.org/resource/22988.aspx)).
- Lautze, S. and Raven-Roberts, A. (2003)** *Food security in complex emergencies: building policy frameworks to address longer-term programming challenges*. Rome: FAO. ([www.alnap.org/resource/7785.aspx](http://www.alnap.org/resource/7785.aspx)).
- Lawday, A. (2014)** *Evaluation of HelpAge International's programme "assistance to specific vulnerable groups affected by the Syrian crisis" (2013-2014)*. London: Andrew Lawday Evaluations. ([www.alnap.org/resource/12857.aspx](http://www.alnap.org/resource/12857.aspx)).
- Lawday, A. (2015)** *Inter-agency humanitarian evaluation (IAHE) of the response to the Central African Republic's crisis 2013-2015*. Washington: Konterra Group. ([www.alnap.org/resource/22989.aspx](http://www.alnap.org/resource/22989.aspx)).
- Laybourn, C. and Obrecht, A. (2014)** *DEC accountability self-assessment validation 2013/14*. London: One World Trust. ([www.alnap.org/resource/22990.aspx](http://www.alnap.org/resource/22990.aspx)).
- Leader, N. (2000)** *The politics of principle: the principles of humanitarian action in practice*. London: ODI/HPG. ([www.alnap.org/resource/8139.aspx](http://www.alnap.org/resource/8139.aspx)).
- Lee, R. M. (2000)** *Unobtrusive methods in social research*. Buckingham: Open University Press. ([www.alnap.org/resource/22991.aspx](http://www.alnap.org/resource/22991.aspx)).
- Lemeshow, S. and Taber, S. (1991)** 'Lot quality assurance sampling: single-and double-sampling plans'. *World Health Stat Q*, 44(3): 115–132. ([www.alnap.org/resource/22992.aspx](http://www.alnap.org/resource/22992.aspx)).
- Leturque, H., Beaujeu, R., Majeed, Y. and Saleem, S. (2012)** *Evaluation of the CBHA early recovery programme in Pakistan*. London: CBHA. ([www.alnap.org/resource/6370.aspx](http://www.alnap.org/resource/6370.aspx)).
- Lieberson, S. (1991)** 'Small N's and big conclusions: An examination of the reasoning in comparative studies based on a small number of cases'. *Social Forces*, 70(2): 307–320. ([www.alnap.org/resource/22994.aspx](http://www.alnap.org/resource/22994.aspx)).
- Lijmer, J. G., Mol, B. W., Heisterkamp, S., Bossel, G. J., Prins, M. H. van der Meulen, J. H. P. and Bossuyt, P. M. M. (1999)** 'Empirical evidence of design-related bias in studies of diagnostic tests'. *JAMA*, 282(11): 1061–1066. ([www.alnap.org/resource/22995.aspx](http://www.alnap.org/resource/22995.aspx)).

- Lindgren, D., Matondang, M. and Putri, D. (2005)** *Tsunami relief study: Effectiveness of the tsunami relief effort*. London: TEC. ([www.alnap.org/resource/22997.aspx](http://www.alnap.org/resource/22997.aspx)).
- Lister, S., Anderson, S., Mokbel Genequand, M., Gordon, A., Sandford, J. and Turner, S. (2011)** *WFP's school feeding policy: a policy evaluation. Vol. 1 full report*. Rome: WFP. ([www.alnap.org/resource/22998.aspx](http://www.alnap.org/resource/22998.aspx)).
- Longo, M., Canetti, D. and Hite-Rubin, N. (2014)** 'A checkpoint effect? Evidence from a natural experiment on travel restrictions in the West Bank'. *American Journal of Political Science*, 58(4): 1006-1023. ([www.alnap.org/resource/23000.aspx](http://www.alnap.org/resource/23000.aspx)).
- Lysy, C. (2014)** *More data, shorter reports: three ways to respond*. Fresh Spectrum ([www.alnap.org/resource/23001.aspx](http://www.alnap.org/resource/23001.aspx)).
- MacDonald, B. and Gedeon, H. (2012)** *Banking with mobile phone: A report on a T-cash pilot project*, Baltimore, MD: CRS. ([www.alnap.org/resource/9882.aspx](http://www.alnap.org/resource/9882.aspx)).
- Mackay, K., Clark, M., Sartorius, R., Bamberger, M. (2004)** *Monitoring & evaluation: Some tools, methods & approaches*. Washington, DC: World Bank. ([www.alnap.org/resource/23003.aspx](http://www.alnap.org/resource/23003.aspx)).
- Mahoney, J. (2000)** 'Strategies of causal inference in small-N analysis'. *Sociological Methods & Research*, 28(4): 387–424. ([www.alnap.org/resource/23004.aspx](http://www.alnap.org/resource/23004.aspx)).
- Mahoney, J. and Goertz, G. (2006)** 'A tale of two cultures: Contrasting quantitative and qualitative research'. *Political Analysis*, 14(3): 227–249. ([www.alnap.org/resource/22637.aspx](http://www.alnap.org/resource/22637.aspx)).
- Majewski, B., Boulet-Desbureau, P., Slezak, M., De Meulder, F. and Wilson, K. (2012)** *Joint evaluation of the global logistics cluster*. ([www.alnap.org/resource/7904.aspx](http://www.alnap.org/resource/7904.aspx)).
- Malki, E. (2008)** *A benchmarking model for measuring the efficiency of a humanitarian aid program: a case study of an international NGO*. Munich: MPRA. ([www.alnap.org/resource/23006.aspx](http://www.alnap.org/resource/23006.aspx)).
- Manesh, A. O., Sheldon, T., A., Pickett, K. E., and Carr-Hill, R. (2008)** 'Accuracy of child morbidity data in demographic and health surveys'. *International Journal of Epidemiology*, 37(1). ([www.alnap.org/resource/8140.aspx](http://www.alnap.org/resource/8140.aspx)).

- Manning, R. (2009)** *The quality of DFID's evaluations and evaluation management systems: How do they compare with other agencies?* London: IACDI. ([www.alnap.org/resource/23007.aspx](http://www.alnap.org/resource/23007.aspx)).
- UN OIOS. (2014)** *Inspection and evaluation manual*. New York: UN OIOS. ([www.alnap.org/resource/19243.aspx](http://www.alnap.org/resource/19243.aspx)).
- Marshall, M. N. (1996)** 'Sampling for qualitative research'. *Family Practice*, 13(6): 522–526. ([www.alnap.org/resource/23008.aspx](http://www.alnap.org/resource/23008.aspx)).
- Martin, E., Petty, C. and Acidri, J. (2009)** *Livelihoods in crisis: a longitudinal study in Pader, Uganda*. London: HPG/ODI. ([www.alnap.org/resource/23010.aspx](http://www.alnap.org/resource/23010.aspx)).
- Martinez, A. (2009)** *A participatory programme review of the international federation of the Red Cross and Red Crescent societies response to the Peru earthquake of 15th August 2007*. Geneva: IFRC. ([www.alnap.org/resource/8141.aspx](http://www.alnap.org/resource/8141.aspx)).
- Martinez, D. E. (2011)** *The logical framework approach in non-governmental organizations*. Edmonton: University of Alberta. ([www.alnap.org/resource/23011.aspx](http://www.alnap.org/resource/23011.aspx)).
- Mays, N. and Pope, C. (2000)** 'Assessing quality in qualitative research'. *BMJ*. ([www.alnap.org/resource/8142.aspx](http://www.alnap.org/resource/8142.aspx)).
- Maystadt, J. F. (2011)** *Poverty reduction in a refugee-hosting economy: A natural experiment*. Washington, DC: IFPRI. ([www.alnap.org/resource/8143.aspx](http://www.alnap.org/resource/8143.aspx)).
- Mazurana, D., Benelli, P., Gupta, H. and Walker, P. (2011)** *Sex and age matter: Improving humanitarian response in emergencies*. Atlanta, GA: CARE International. ([www.alnap.org/resource/8144.aspx](http://www.alnap.org/resource/8144.aspx)).
- McDonald, B. and Rogers, P. (2014)** *Interviewing*. Florence: UNICEF. ([www.alnap.org/resource/23012.aspx](http://www.alnap.org/resource/23012.aspx)).
- McGearty, S., O'Hagan, P. and Montinard, M. (2012)** *An independent final evaluation of the action of churches together alliance Haiti appeal HTI-101 (Jan 2010 – Dec 2011)*. London: Channel Research. ([www.alnap.org/resource/6339.aspx](http://www.alnap.org/resource/6339.aspx)).
- McRAM Team. (2009)** *Multi-cluster rapid assessment mechanism (McRAM) in Pakistan*. Islamabad: McRAM. ([www.alnap.org/resource/8145.aspx](http://www.alnap.org/resource/8145.aspx)).

**Hanif, A. (2008)** *HAP initiatives in Concern Worldwide: A case study of Bangladesh (2008)*. London: Concern Worldwide. ([www.alnap.org/resource/23013.aspx](http://www.alnap.org/resource/23013.aspx)).

**Mendoza, U. and Thomas, V. (2009)** *UNHCR's AGDM evaluation: A participatory evaluation of AGDM results in four Colombian communities*. Geneva: UNHCR. ([www.alnap.org/node/24503.aspx](http://www.alnap.org/node/24503.aspx)).

**Meng, X. and Qian, N. (2009)** *The long term consequences of famine on survivors: Evidence from a unique natural experiment using China's great famine*. Cambridge, MA: NBER. ([www.alnap.org/resource/23015.aspx](http://www.alnap.org/resource/23015.aspx)).

**Meyer, A.-M. (2007)** *The causes of malnutrition in children under 3 in the Somali Region of Ethiopia related to household caring practices*. London: Save the Children. ([www.alnap.org/resource/3591.aspx](http://www.alnap.org/resource/3591.aspx)).

**Mezuk, B., Larkin, G. L., Prescott, M. R., Tracy, M., Vlahov, D., Tardiff, K. and Galea, S. (2009)** 'The influence of a major disaster on suicide risk in the population'. *Journal of Traumatic Stress*, 22(6): 481-488. ([www.alnap.org/resource/23017.aspx](http://www.alnap.org/resource/23017.aspx)).

**Miles, M. B., Huberman, A. M. and Saldaña, J. (2013)** *Qualitative data analysis: A methods sourcebook*. 3rd ed. Thousand Oaks, CA: SAGE. ([www.alnap.org/resource/23018.aspx](http://www.alnap.org/resource/23018.aspx)).

**Miller, J. and Stolz, K. (2008)** *Moving beyond rhetoric: Consultation and participation with populations displaced by conflict or natural disasters*. Washington, DC: The Brookings Institution. ([www.alnap.org/resource/8751.aspx](http://www.alnap.org/resource/8751.aspx)).

**Milojevic, A., Armstrong, B., Hashizume, M., McAllister, K., Faruque, A., Yunus, M., Kim Streatfield, P. Moji, K. and Wilkinson, P. (2012)** 'Health effects of flooding in rural Bangladesh'. *Epidemiology*, 23(1): 107–115. ([www.alnap.org/resource/23019.aspx](http://www.alnap.org/resource/23019.aspx)).

**Milojevic, A., Armstrong, B., Kovats, S., Butler, B., Hayes, E., Leonardi, G., Murray, V. and Wilkinson, P. (2011)** 'Long-term effects of flooding on mortality in England and Wales, 1994-2005: controlled interrupted time-series analysis'. *Environmental Health*. ([www.alnap.org/resource/23020.aspx](http://www.alnap.org/resource/23020.aspx)).

**Miranda, R., (2009)** *Eva the Evaluator*. Learningham Press. ([www.alnap.org/resource/23028.aspx](http://www.alnap.org/resource/23028.aspx)).

**Mirano, S. (2012)** *Learning from the Urban Transitional Shelter Response in Haiti*. Baltimore, MD: CRS. ([www.alnap.org/resource/6539.aspx](http://www.alnap.org/resource/6539.aspx)).

**Molund, S. and Schill, G. (2007)** *Looking back, moving forward*. 2nd ed. Stockholm: SIDA. ([www.alnap.org/resource/7902.aspx](http://www.alnap.org/resource/7902.aspx)).

**Molyneux, S., Mulupi, S., Mbaabu, L. and Marsh, V. (2012)** 'Benefits and payments for research participants: Experiences and views from a research centre on the Kenyan coast'. *BMC Medical Ethics*, 13(1). ([www.alnap.org/resource/8147.aspx](http://www.alnap.org/resource/8147.aspx)).

**Montbiot, E. (2006)** *World Vision: Lessons by Sectors*. Monrovia, CA: World Vision. ([www.alnap.org/resource/23029.aspx](http://www.alnap.org/resource/23029.aspx)).

**Morán, J. L. Á. (2012)** *External evaluation: Emergency nutrition programme in Sindh Province, Pakistan*. London: Action Against Hunger. ([www.alnap.org/resource/20934.aspx](http://www.alnap.org/resource/20934.aspx)).

**Morel, D. and Hagens, C. (2012)** *Monitoring, evaluation, accountability and learning in emergencies - a resource pack for simple and strong MEAL*. Baltimore, MD: CRS. ([www.alnap.org/resource/9200.aspx](http://www.alnap.org/resource/9200.aspx)).

**Morgan, D. L. (1997)** *Focus groups as qualitative research*. 2nd ed. Thousand Oaks, CA: SAGE. ([www.alnap.org/resource/23030.aspx](http://www.alnap.org/resource/23030.aspx)).

**Morinière, L. (2011a)** *Evaluation of the ERRF component of the Haiti emergency response (ERF) fund*. New York: OCHA. ([www.alnap.org/resource/6054.aspx](http://www.alnap.org/resource/6054.aspx)).

**Morinière, L. (2011b)** *External evaluation of the Haiti emergency relief and response fund (ERRF), 2008-2011*. New York: OCHA. ([www.alnap.org/resource/8149.aspx](http://www.alnap.org/resource/8149.aspx)).

**Morra-Imas, L. G. and Rist, R. C. (2009)** *The road to results: Designing and conducting effective development evaluations*. Washington, DC: World Bank. ([www.alnap.org/resource/8470.aspx](http://www.alnap.org/resource/8470.aspx)).

**Morris, M. F. and Shaughnessy, D. E. (2007)** *Final evaluation report: Emergency capacity building project*. Atlanta, GA: CARE International. ([www.alnap.org/resource/10636.aspx](http://www.alnap.org/resource/10636.aspx)).

**Morris, T., Steen, N. and Canteli, C. (2013)** *Synthesis of mixed method impact evaluations of the contribution of food assistance to durable solutions in protracted refugee situations*. Geneva: UNHCR. ([www.alnap.org/resource/19902.aspx](http://www.alnap.org/resource/19902.aspx)).

**Moscoe, E., Bor, J. and Bärnighausen, T. (2015)** 'Regression discontinuity designs are underutilized in medicine, epidemiology, and public health: a review of current and best practice'. *Journal of Clinical Epidemiology*, 68(2): 132–143. (www.alnap.org/resource/23031.aspx).

**Mowjee, T., Fleming, D. and Toft, E. (2015)** *Evaluation of the strategy for Danish humanitarian action 2010-2015*. Copenhagen: DANIDA. (www.alnap.org/resource/20738.aspx).

**MSF, 2010.** *Summary of findings of emergency response evaluations*. Geneva: MSF. (www.alnap.org/resource/23032.aspx).

**MSF. (2012)** *A handbook for initiating, managing and conducting evaluations in MSF*. Geneva: MSF. (www.alnap.org/resource/10004.aspx).

**Must, A., Phillips, S., Naumova, E., Blum, M., Harris, S., Dawson-Hughes, B. and Rand, W. (2002)** 'Recall of early menstrual history and menarcheal body size: After 30 years, how well do women remember?' *American Journal of Epidemiology*, 155(7): 672–679. (www.alnap.org/resource/8150.aspx).

**La Pelle, N. R. (2004)** 'Simplifying qualitative data analysis using general purpose software tools'. *Field Methods*, 16(1): 85–108. (www.alnap.org/resource/23033.aspx).

**Naidu, V. J. (2010)** *A thematic evaluation of livelihood and community based disaster preparedness projects*. Atlanta, GA: CARE International. (www.alnap.org/resource/10252.aspx).

**NAO. (2000)** *A practical guide to sampling*. London: NAO. (www.alnap.org/resource/10228.aspx).

**NORAD. (1998)** *Results management in Norwegian development cooperation : A practical guide*. Oslo: NORAD. (www.alnap.org/resource/8052.aspx).

**NORAD. (1999)** *Logical framework approach: handbook for objectives-oriented planning*. Oslo: NORAD. (www.alnap.org/resource/23034.aspx).

**Norman, B. (2012)** *Monitoring and accountability practices for remotely managed projects implemented in volatile operating environments*. Teddington: Tearfund. (www.alnap.org/resource/7956.aspx).

**NRC. (2013)** *Evaluation of five humanitarian programmes of the Norwegian Refugee Council (NRC) and of the standby roster NORCAP. Case country report – Somalia.* Oslo: NRC. ([www.alnap.org/resource/12227.aspx](http://www.alnap.org/resource/12227.aspx)).

**O'Hagan, P., Love, K. and Rouse, A. (2010)** *An independent joint evaluation of the Haiti earthquake humanitarian response.* Atlanta, GA: CARE International. ([www.alnap.org/resource/5969.aspx](http://www.alnap.org/resource/5969.aspx)).

**O'Neil, G. (2012)** *7 new ways to present evaluation findings.* 3-5 October, Helsinki. ([www.alnap.org/resource/23035.aspx](http://www.alnap.org/resource/23035.aspx)).

**Oakden, J. (2013a)** *Evaluation rubrics: how to ensure transparent and clear assessment that respects diverse lines of evidence.* Melbourne: Better Evaluation. ([www.alnap.org/resource/23036.aspx](http://www.alnap.org/resource/23036.aspx)).

**Oakden, J., 2013b.** *Guest blog: Why rubrics are useful in evaluations.* Melbourne: Better Evaluation. ([www.alnap.org/resource/23037.aspx](http://www.alnap.org/resource/23037.aspx)).

**Obrecht, A., Laybourne, C., Hammer, M. and Ray, S. (2012)** *The 2011/12 DEC accountability framework assessment: findings from the external evaluation and peer review process.* London: One World Trust. ([www.alnap.org/resource/23038.aspx](http://www.alnap.org/resource/23038.aspx)).

**OCHA. (2004)** *Guiding principles on internal displacement.* New York: OCHA. ([www.alnap.org/resource/8054.aspx](http://www.alnap.org/resource/8054.aspx)).

**OCHA. (2009)** 'OCHA on message : Humanitarian access case study'. *OCHA annual report 2008.* New York: OCHA. ([www.alnap.org/resource/11252.aspx](http://www.alnap.org/resource/11252.aspx)).

**OCHA. (2011)** *Inter-agency RTE for the 2011 Pakistan floods.* New York: OCHA. ([www.alnap.org/resource/23040.aspx](http://www.alnap.org/resource/23040.aspx)).

**OCHA. (2012)** *OCHA on message: humanitarian principles.* New York: OCHA. ([www.alnap.org/resource/19270.aspx](http://www.alnap.org/resource/19270.aspx)).

**OCHA. (2013)** *United Nations Disaster Assessment and Coordination UNDAC field handbook.* New York: OCHA. ([www.alnap.org/resource/23041.aspx](http://www.alnap.org/resource/23041.aspx)).

**ODI. (2013)** *Research and policy in development (RAPID).* London: ODI. ([www.alnap.org/resource/23048.aspx](http://www.alnap.org/resource/23048.aspx)).

**OECD. (2010)** *Managing Joint Evaluations*. Paris: OECD.  
([www.alnap.org/resource/23039.aspx](http://www.alnap.org/resource/23039.aspx)).

**OECD/DAC. (1991)** *Principles for evaluation of development assistance*. Paris: OECD/DAC. ([www.alnap.org/resource/20830.aspx](http://www.alnap.org/resource/20830.aspx)).

**OECD/DAC. (1999)** *Guidance for evaluating humanitarian assistance in complex emergencies*. Paris: OECD/DAC. ([www.alnap.org/resource/8221.aspx](http://www.alnap.org/resource/8221.aspx)).

**OECD/DAC. (2010)** *Glossary of key terms in evaluation and results based management*. Paris: OECD/DAC. ([www.alnap.org/resource/8489.aspx](http://www.alnap.org/resource/8489.aspx)).

**OECD/DAC. (2012)** *Evaluating peacebuilding activities in settings of conflict and fragility: Improving learning for results*. Paris: OECD/DAC. ([www.alnap.org/resource/8057.aspx](http://www.alnap.org/resource/8057.aspx)).

**OIOS. (2014)** *Inspection and evaluation manual*. New York: OIOS.  
([www.alnap.org/resource/19243.aspx](http://www.alnap.org/resource/19243.aspx)).

**Ojha, G. P. (2010)** *Appreciative inquiry approach to evaluation practices in South Asia*. Iasi: Lumen Publishing House. ([www.alnap.org/resource/23042.aspx](http://www.alnap.org/resource/23042.aspx)).

**Oliver, M. L. (2007)** *CARE's humanitarian operations: Review of CARE's use of evaluations and after action reviews in decision-making*. Atlanta, GA: CARE International. ([www.alnap.org/resource/8058.aspx](http://www.alnap.org/resource/8058.aspx)).

**Oliver, M. L. (2009)** 'Metaevaluation as a means of examining evaluation influence'. *Journal of MultiDisciplinary Evaluation*, 6(11): 32–37. ([www.alnap.org/resource/8285.aspx](http://www.alnap.org/resource/8285.aspx)).

**Ong, J. C., Flores, J. M. and Combinido, P. (2015)** *Obligated to be grateful*. Woking: Plan International. ([www.alnap.org/resource/20633.aspx](http://www.alnap.org/resource/20633.aspx)).

**Onwuegbuzie, A. J. and Collins, K. M. T. (2007)** 'A typology of mixed methods sampling designs in social science research'. *The Qualitative Report*, 12(2): 281–316. ([www.alnap.org/resource/23044.aspx](http://www.alnap.org/resource/23044.aspx)).

**Onwuegbuzie, A. J. and Leech, N. (2007)** 'A call for qualitative power analyses'. *Quality & Quantity*, 41(1): 105–121. ([www.alnap.org/resource/23045.aspx](http://www.alnap.org/resource/23045.aspx)).

**Ott, E., Krystalli, R. C., Stites, E., Timmins, N., Walden, V., Gillingham, E., Bushby, K. (2015)** *Humanitarian evidence synthesis and communication programme: Abridged inception report*. Oxford: Oxfam and Feinstein International Center. ([www.alnap.org/resource/23046.aspx](http://www.alnap.org/resource/23046.aspx)).

**Otto, R. (2015)** *Story in 5: Ralf Otto, Momologue and Ebaix*. [YouTube] London: ALNAP. ([www.alnap.org/resource/23047.aspx](http://www.alnap.org/resource/23047.aspx)).

**Otto, R., Conrad, C., Brusset, E. and Stone, L. (2006)** *Organisational learning review of Caritas Internationalis' response to the Tsunami emergency*. London: Channel Research. ([www.alnap.org/resource/8059.aspx](http://www.alnap.org/resource/8059.aspx)).

**Oxfam. (2011)** *How are effectiveness reviews carried out?* Oxford: Oxfam. ([www.alnap.org/resource/8473.aspx](http://www.alnap.org/resource/8473.aspx)).

**Oxfam. (2012)** *Project effectiveness reviews: summary of 2011/12 findings and lessons learned*. Oxford: Oxfam. ([www.alnap.org/resource/8472.aspx](http://www.alnap.org/resource/8472.aspx)).

**Page, L., Savage, D. and Torgler, B. (2012)** *Variation in risk seeking behavior in a natural experiment on large losses induced by a natural disaster*. Zürich: CREMA. ([www.alnap.org/resource/23050.aspx](http://www.alnap.org/resource/23050.aspx)).

**Pandya, M., Pathak, V. and Oza, S. (2012)** *Future of M&E in humanitarian sector: Possible list for discussion*. Gujarat: AIDMI. ([www.alnap.org/resource/22894.aspx](http://www.alnap.org/resource/22894.aspx)).

**Pantera, G. (2012)** *AGIRE humanitarian response to the East Africa drought*. Rome: AGIRE. ([www.alnap.org/resource/6316.aspx](http://www.alnap.org/resource/6316.aspx)).

**Patrick, J. (2011)** *Haiti earthquake response. Emerging evaluation lessons*. London: ALNAP. ([www.alnap.org/resource/6125.aspx](http://www.alnap.org/resource/6125.aspx)).

**Patton, M. Q. (1997)** *Utilisation-focused evaluation: the new century text*. Thousand Oaks, CA: SAGE. ([www.alnap.org/resource/11131.aspx](http://www.alnap.org/resource/11131.aspx)).

**Patton, M. Q. (2008)** *Utilization-focused evaluation*. Thousand Oaks, CA: SAGE. ([www.alnap.org/resource/10060.aspx](http://www.alnap.org/resource/10060.aspx)).

**Patton, M. Q. and Cochran, M. (2002)** *A guide to using qualitative research methodology*. Geneva: MSF. ([www.alnap.org/resource/13024.aspx](http://www.alnap.org/resource/13024.aspx)).

- Pawson, R. and Tilley, N. (2004)** *Realist evaluation*. Mount Torrens: Community Matters. ([www.alnap.org/resource/8195.aspx](http://www.alnap.org/resource/8195.aspx)).
- Peersman, G. (2014a)** *Evaluative criteria*. Florence: UNICEF. ([www.alnap.org/resource/23051.aspx](http://www.alnap.org/resource/23051.aspx)).
- Peersman, G. (2014b)** *Overview: Data collection and analysis methods in impact evaluation*. Florence: UNICEF. ([www.alnap.org/resource/22654.aspx](http://www.alnap.org/resource/22654.aspx)).
- Penfold, R. B. and Zhang, F. (2013)** 'Use of interrupted time series analysis in evaluating health care quality improvements'. *Academic Pediatrics*, 13(6): S38–S44. ([www.alnap.org/resource/23052.aspx](http://www.alnap.org/resource/23052.aspx)).
- People in Aid. (2003)** *Code of good practice in the management and support of aid personnel*. London: People in Aid. ([www.alnap.org/resource/8061.aspx](http://www.alnap.org/resource/8061.aspx)).
- Perrin, B. (2012)** *Linking Monitoring and Evaluation to Impact Evaluation*. Washington, DC: InterAction. ([www.alnap.org/resource/19267.aspx](http://www.alnap.org/resource/19267.aspx)).
- Petticrew, M., Cummins, S., Ferrell, C., Findlay, A., Higgins, C. Hoy, C., Kearns, A. and Sparks, L. (2005)** 'Natural experiments: an underused tool for public health?' *Public Health*, 119(9): 751–757. ([www.alnap.org/resource/23053.aspx](http://www.alnap.org/resource/23053.aspx)).
- Pfister, R. (2011)** 'Wardrobe malfunctions and the measurement of internet behaviour'. *Psychology*, 2(3): 266–268. ([www.alnap.org/resource/23054.aspx](http://www.alnap.org/resource/23054.aspx)).
- Picciotto, R. (2012)** 'Experimentalism and development evaluation: Will the bubble burst?' *Evaluation*, 18(2): 213–229. ([www.alnap.org/resource/8062.aspx](http://www.alnap.org/resource/8062.aspx)).
- Rawlins, R., Pimkina, S., Barrett, C. B., Pedersen, S. and Wydick, B. (2013)** 'Got milk? The impact of Heifer International's livestock donation programs in Rwanda'. *Food Policy*, 44(2): 202–213. ([www.alnap.org/resource/23055.aspx](http://www.alnap.org/resource/23055.aspx)).
- Polastro, R. (2014)** *Evaluating humanitarian action in real time: Recent practices, challenges, and innovations*. Sheffield: IOD Parc. ([www.alnap.org/resource/12928.aspx](http://www.alnap.org/resource/12928.aspx)).
- Polastro, R., Khalif, M. A., van Ebyen, M., Posada, S., Salah, A., Steen, N. and Toft, E. (2011)** *IASC evaluation of the humanitarian response in South Central Somalia 2005-2010*. Madrid: DARA. ([www.alnap.org/resource/9301.aspx](http://www.alnap.org/resource/9301.aspx)).

**Polastro, R., Nagrah, A., Steen, N. and Zafar, F. (2011)** *Inter-agency real time evaluation of the humanitarian response to Pakistan's 2010 Flood Crisis*. Madrid: DARA. ([www.alnap.org/resource/6087.aspx](http://www.alnap.org/resource/6087.aspx)).

**Poulos, C., Pattanayak, S. K. and Jones, K. (2006)** *A guide to water and sanitation sector impact evaluations*. Washington, DC: World Bank. ([www.alnap.org/resource/8063.aspx](http://www.alnap.org/resource/8063.aspx)).

**Poulsen, L., Stacy, R., Bell, L. and Kumar Range, S. (2009)** *Joint thematic evaluation of FAO and WFP support to information systems for food security*. Rome: WFP. ([www.alnap.org/resource/5845.aspx](http://www.alnap.org/resource/5845.aspx)).

**Preskill, H. S. and Catsambas, T. T. (2006)** *Reframing evaluation through appreciative inquiry*. Thousand Oaks, CA: SAGE. ([www.alnap.org/resource/8224.aspx](http://www.alnap.org/resource/8224.aspx)).

**Proudlock, K., Ramalingam, B. and Sandison, P. (2009)** 'Improving humanitarian impact assessment: bridging theory and practice', in *ALNAP's 8th review of humanitarian action: Performance, Impact and Innovation*. ALNAP Review. London: ALNAP. ([www.alnap.org/resource/5663.aspx](http://www.alnap.org/resource/5663.aspx)).

**Puoane, T., Alexander, L. and Hutton, B. (2011)** *The revolving door: child malnutrition in Mount Frere, Eastern Cape Province of South Africa*. Cape Town: University of the Western Cape. ([www.alnap.org/resource/23056.aspx](http://www.alnap.org/resource/23056.aspx)).

**Puri, J., Aladysheva, A., Iversen, V., Ghorpade, Y. and Brück, T. (2014)** *What methods may be used in impact evaluations of humanitarian assistance?* New Delhi: 3ie. ([www.alnap.org/resource/19288.aspx](http://www.alnap.org/resource/19288.aspx)).

**Qasim, Q., 3le and Stevens, K. (2014)** *Regression Discontinuity*. Melbourne: Better Evaluation. ([www.alnap.org/resource/23057.aspx](http://www.alnap.org/resource/23057.aspx)).

**Rabbani, M., Prakash, V. A. and Sulaiman, M. (2006)** *Impact assessment of CFPR/TUP: a descriptive analysis based on 2002-2005 panel data*. Dhaka: BRAC. ([www.alnap.org/resource/23058.aspx](http://www.alnap.org/resource/23058.aspx)).

**Rademaker, L. L., Grace, E. J. and Curda, S. K. (2012)** 'Using computer-assisted qualitative data analysis software (CAQDAS) to re-examine traditionally analyzed data: Expanding our understanding of the data and of ourselves as scholars'. *The Qualitative Report*, 17(22): 1–11. ([www.alnap.org/resource/23059.aspx](http://www.alnap.org/resource/23059.aspx)).

**Ramalingam, B. (2011)** *Learning how to learn: eight lessons for impact evaluations that make a difference*. London: ODI. ([www.alnap.org/resource/8458.aspx](http://www.alnap.org/resource/8458.aspx)).

**Ramalingam, B. and Mitchell, J. (2014)** *Responding to changing needs?* London: ALNAP. ([www.alnap.org/resource/19246.aspx](http://www.alnap.org/resource/19246.aspx)).

**Raphael, K. (1987)** 'Recall bias: A proposal for assessment and control'. *International Journal of Epidemiology*, 16(2): 167–170. ([www.alnap.org/resource/23060.aspx](http://www.alnap.org/resource/23060.aspx)).

**Raworth, K. (2016)** *Reviewing the existing literature*. Oxford: Oxfam. ([www.alnap.org/resource/23049.aspx](http://www.alnap.org/resource/23049.aspx)).

**Reddit Inc. (2013)** *Reddit: the front page of the internet*. ([www.alnap.org/resource/23061.aspx](http://www.alnap.org/resource/23061.aspx)).

**Reid, E. and Novak, P. (1975)** 'Personal space: An unobtrusive measures study'. *Bulletin of the Psychonomic Society*, 5(3): 265–266. ([www.alnap.org/resource/23062.aspx](http://www.alnap.org/resource/23062.aspx)).

**Reid, S., Stibbe, D. and Lowery, C. (2010)** *Review of the global education cluster co-leadership arrangement*. Oxford: The Partnering Initiative. ([www.alnap.org/resource/6354.aspx](http://www.alnap.org/resource/6354.aspx)).

**Remler, D. K. and Van Ryzin, G. G. (2014)** *Research methods in practice: Strategies for description and causation*. 2nd ed. Thousand Oaks, CA: SAGE. ([www.alnap.org/resource/23063.aspx](http://www.alnap.org/resource/23063.aspx)).

**Ridde, V., Haddad, S. and Heinmüller, R. (2013)** 'Improving equity by removing healthcare fees for children in Burkina Faso'. *Journal of Epidemiology and Community Health*. 2013(0): 1-7. ([www.alnap.org/resource/23064.aspx](http://www.alnap.org/resource/23064.aspx)).

**Riddell, R. C. (2009)** *The quality of DFID's evaluation reports and assurance systems: A report for IACDI based on the quality review undertaken by consultants Burt Perrin and Richard Manning*. London: IACDI. ([www.alnap.org/resource/8496.aspx](http://www.alnap.org/resource/8496.aspx)).

**Riessman, C. K. (1977)** *The effect of interviewer status and respondent sex on symptom reporting*. ([www.alnap.org/resource/8064.aspx](http://www.alnap.org/resource/8064.aspx)).

- Rist, R. C., Boily, M.-H. and Martin, F. (eds) (2011)** *Influencing change: building evaluation capacity to strengthen governance*. Washington, DC: World Bank. ([www.alnap.org/resource/23065.aspx](http://www.alnap.org/resource/23065.aspx)).
- Ritter, T. and Estoesta, J. (2015)** *Ethical challenges in the evaluation of a Pacific multi-country training project by an Australian NGO*. Deakin, ACT: ACFID. ([www.alnap.org/resource/23066.aspx](http://www.alnap.org/resource/23066.aspx)).
- Robert, S. and Valon, A. (2014)** *Quality and Accountability for Project Cycle Management*.
- Roche, C. (1999)** *Impact assessment for development agencies: Learning to value change*. Oxford: Oxfam. ([www.alnap.org/resource/8065.aspx](http://www.alnap.org/resource/8065.aspx)).
- Roche, C. (2000)** *Impact assessment: Seeing the wood and the trees*. Oxford: Oxfam. ([www.alnap.org/resource/11135.aspx](http://www.alnap.org/resource/11135.aspx)).
- Rogers, P. (2012)** *Introduction to impact evaluation*. Washington, DC: InterAction. ([www.alnap.org/resource/6387.aspx](http://www.alnap.org/resource/6387.aspx)).
- Rogers, P. (2013)** *52 weeks of BetterEvaluation: Week 11: Using rubrics*. Melbourne: Better Evaluation. ([www.alnap.org/resource/23067.aspx](http://www.alnap.org/resource/23067.aspx)).
- Rogers, P. (2014a)** *Overview of impact evaluation*. Florence: UNICEF. ([www.alnap.org/resource/22671.aspx](http://www.alnap.org/resource/22671.aspx)).
- Rogers, P. (2014b)** *Overview: Strategies for causal attribution*. Florence: UNICEF. ([www.alnap.org/resource/22674.aspx](http://www.alnap.org/resource/22674.aspx)).
- Rogers, P. (2014c)** *Theory of change*. Florence: UNICEF. ([www.alnap.org/resource/22673.aspx](http://www.alnap.org/resource/22673.aspx)).
- Rosenberg, L. J., Posner, L. D. and Hanley, E. J. (1970)** *Project evaluation and the project appraisal reporting system: Volume one: Summary*. Washington: USAID. ([www.alnap.org/resource/23068.aspx](http://www.alnap.org/resource/23068.aspx)).
- Rouse, A. (2012)** *After action review (AAR) report*. Atlanta, GA: CARE International. ([www.alnap.org/resource/8495.aspx](http://www.alnap.org/resource/8495.aspx)).

**Ryan, M. E. (2009)** 'Making visible the coding process: Using qualitative data software in a post-structural study'. *Issues in Educational Research*, 19(2): 142–161. ([www.alnap.org/resource/23069.aspx](http://www.alnap.org/resource/23069.aspx)).

**Sackett, D. L. (1979)** 'Bias in analytic research'. *Journal of Chronic Diseases*, 32(1–2): 51–63. ([www.alnap.org/resource/23070.aspx](http://www.alnap.org/resource/23070.aspx)).

**Sackett, D. L. (2007)** 'Commentary: Measuring the success of blinding in RCTs: don't, must, can't or needn't?' *International Journal of Epidemiology*, 36(3): 664–665. ([www.alnap.org/resource/23071.aspx](http://www.alnap.org/resource/23071.aspx)).

**Salama, P., Assefa, F., Talley, L., Spiegel, P., van der Veen, A. and Gotway, C. A. (2001)** 'Malnutrition, measles, mortality, and the humanitarian response during a famine in Ethiopia'. *JAMA*, 286(5): 563–571. ([www.alnap.org/resource/23072.aspx](http://www.alnap.org/resource/23072.aspx)).

**Sandison, P. (2006)** 'The utilisation of evaluations', in *ALNAP review of humanitarian action in 2005: Evaluation utilisation*. ALNAP Review. London: ALNAP. ([www.alnap.org/resource/5225.aspx](http://www.alnap.org/resource/5225.aspx)).

**Sandström, S. and Tchatchua, L. (2010)** 'Do cash transfers improve food security in emergencies? Evidence from Sri Lanka', in Omamo, S. W., Gentilini, U. and Sandström, S. (eds), *Revolution: From food aid to food assistance: Innovations in overcoming hunger*. Rome: WFP. ([www.alnap.org/resource/8077.aspx](http://www.alnap.org/resource/8077.aspx)).

**Saunders, M. (2011)** *Evaluation use and usability: theoretical foundations and current state of international thinking*. 8 July, Gdansk. ([www.alnap.org/resource/23073.aspx](http://www.alnap.org/resource/23073.aspx)).

**Savage, K., Delesgues, L., Martin, E. and Ulfat, G. P. (2007)** *Corruption perceptions and risks in humanitarian assistance: An Afghanistan case study*. London: HPG/ODI. ([www.alnap.org/resource/22801.aspx](http://www.alnap.org/resource/22801.aspx)).

**Sawada, Y. and Shimizutani, S. (2008)** 'How do people cope with natural disasters? Evidence from the great Hanshin-Awaji (Kobe) earthquake in 1995'. *Journal of Money, Credit and Banking*, 40(2-3): 463–488. ([www.alnap.org/resource/23074.aspx](http://www.alnap.org/resource/23074.aspx)).

**Schmidt, W.-P., Arnold, B. F., Boisson, S., Genser, B., Luby, S. P., Barreto, M. L., Clasen, T. and Cairncross, S. (2011)** 'Epidemiological methods in diarrhoea studies—an update'. *International Journal of Epidemiology*, 40(6): 1678–1692. ([www.alnap.org/resource/23075.aspx](http://www.alnap.org/resource/23075.aspx)).

**Schulz, K. F., Chalmers, I., Hayes, R. J. and Altman, D. G. (1995)** 'Empirical evidence of bias: Dimensions of methodological quality associated with estimates of treatment effects in controlled trials'. *JAMA*, 273(5): 408–412. ([www.alnap.org/resource/23076.aspx](http://www.alnap.org/resource/23076.aspx)).

**Schwartz, L., Sinding, C., Hunt, M., Elit, L., Redwood-Campbell, L., Adelson, N., Luther, L. Ranford, J. and DeLaat, S. (2010)** 'Ethics in humanitarian aid work: Learning from the narratives of humanitarian health workers'. *AJOB Primary Research*, 1(3): 45–54. ([www.alnap.org/resource/23077.aspx](http://www.alnap.org/resource/23077.aspx)).

**Scriven, M. (1991)** *Evaluation thesaurus*. Newbury Park, CA: SAGE. ([www.alnap.org/resource/8480.aspx](http://www.alnap.org/resource/8480.aspx)).

**Scriven, M. (2008)** 'A summative evaluation of RCT methodology: and an alternative approach to causal research'. *Journal of MultiDisciplinary Evaluation*, 5(9): 14. ([www.alnap.org/resource/8153.aspx](http://www.alnap.org/resource/8153.aspx)).

**Scriven, M. (2011)** *Evaluating evaluations: a meta-evaluation checklist*. ([www.alnap.org/resource/8154.aspx](http://www.alnap.org/resource/8154.aspx)).

**Seebregts, C. J., Zwarenstein, M., Mathews, C., Fairall, L., Flisher, A. J., Seebregts, C., Mukoma, W. and Klepp, K. I. (2009)** 'Handheld computers for survey and trial data collection in resource-poor settings: Development and evaluation of PDACT, a Palm Pilot interviewing system'. *International journal of medical informatics*. ([www.alnap.org/resource/8155.aspx](http://www.alnap.org/resource/8155.aspx)).

**Serrat, O. (2009)** *The most significant change technique*. Manila: Asian Development Bank. ([www.alnap.org/resource/22802.aspx](http://www.alnap.org/resource/22802.aspx)).

**Shackman, G. (2001)** *Sample size and design effect*, Albany: American Statistical Association. ([www.alnap.org/resource/23080.aspx](http://www.alnap.org/resource/23080.aspx)).

**Shadish, W. R., Cook, T. D. and Campbell, D. T. (2002)** 'Experiments and generalized causal inference and a critical assessment of our assumptions', in *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin Company. ([www.alnap.org/resource/22889.aspx](http://www.alnap.org/resource/22889.aspx)).

**Shah, R. (2014)** *Evaluation of the Norwegian Refugee Council's Palestine education programme 2010-2014*. Oslo: NRC. ([www.alnap.org/resource/12820.aspx](http://www.alnap.org/resource/12820.aspx)).

**Shanks, L., Ariti, C., Siddiqui, M. R., Pintaldi, G., Venis, S. de Jong, K. and Denault, M. (2013)** 'Counselling in humanitarian settings: a retrospective analysis of 18 individual-focused non-specialised counselling programmes'. *Conflict and Health*, 7(19). ([www.alnap.org/resource/23081.aspx](http://www.alnap.org/resource/23081.aspx)).

**Shannon, H. S., Hutson, R., Kolbe, A., Stringer, B., Haines, T., Nelson, N., Yang, L., Reilly, A., Hardin, J. and Hartley, D. (2012)** 'Choosing a survey sample when data on the population are limited'. *Emerging Themes in Epidemiology*. ([www.alnap.org/resource/8156.aspx](http://www.alnap.org/resource/8156.aspx)).

**Shaw, I., Greene, J. C. and Mark, M. M. (2006)** *The SAGE handbook of evaluation: policies, programs and practices*. London: SAGE. ([www.alnap.org/resource/23082.aspx](http://www.alnap.org/resource/23082.aspx)).

**Shirima, K., Mukasa, O., Armstrong Schellenberg, J., Manzi, F., John, D., Mushi, A., Mrisho, M., Tanner, M., Mshinda, H. and Schellenberg, D. (2007)** 'The use of personal digital assistants for data entry at the point of collection in a large household survey in southern Tanzania'. *Emerging themes in epidemiology 2007*. ([www.alnap.org/resource/8157.aspx](http://www.alnap.org/resource/8157.aspx)).

**SIDA. (2011)** *Strategic Evaluation Plan 2012*. Stockholm: SIDA. ([www.alnap.org/resource/23083.aspx](http://www.alnap.org/resource/23083.aspx)).

**Sida, L. (2013)** *Syria crisis: Evaluators learning exchange on 13 September 2013*. London: ALNAP. ([www.alnap.org/resource/9300.aspx](http://www.alnap.org/resource/9300.aspx)).

**Sida, L. (2014)** *Remote monitoring, evaluation and accountability in the Syria response*. London: ALNAP. ([www.alnap.org/resource/12803.aspx](http://www.alnap.org/resource/12803.aspx)).

**Sida, L., Gray, B. and Asmare, E. (2012)** *IASC real time evaluation (IASC RTE) of the humanitarian response to the Horn of Africa drought crisis – Ethiopia*. Geneva: IASC. ([www.alnap.org/resource/7505.aspx](http://www.alnap.org/resource/7505.aspx)).

**Silver, C. and Lewins, A. (2014)** *Using software in qualitative research: a step by step guide*. 2nd ed. Los Angeles, CA: SAGE. ([www.alnap.org/resource/23084.aspx](http://www.alnap.org/resource/23084.aspx)).

**Slim, H. and Trombetta, L. (2014)** *Syria crisis common context analysis*, New York: OCHA. ([www.alnap.org/resource/12718.aspx](http://www.alnap.org/resource/12718.aspx)).

**DANIDA. (2015)** *Evaluation of the Danish engagement in Palestine*. Copenhagen: DANIDA. ([www.alnap.org/resource/23085.aspx](http://www.alnap.org/resource/23085.aspx)).

**SLRC. (2012)** *Making systematic reviews work for international development*. London: SLRC/ODI. ([www.alnap.org/resource/2079.aspx](http://www.alnap.org/resource/2079.aspx)).

**Small, M .L. (2011)** 'How to conduct a mixed methods study: Recent trends in a rapidly growing literature'. *Annual Review of Sociology*, 37(1): 57–86. ([www.alnap.org/resource/23086.aspx](http://www.alnap.org/resource/23086.aspx)).

**SMART, 2012.** *Sampling methods and sample size calculation for the SMART methodology*. Toronto: SMART. ([www.alnap.org/resource/23087.aspx](http://www.alnap.org/resource/23087.aspx)).

**Smutylo, T. (2001)** *Crouching impact, hidden attribution: Overcoming threats to learning in development programs*. Ottawa: International Development Research Centre. ([www.alnap.org/resource/19253.aspx](http://www.alnap.org/resource/19253.aspx)).

**Sphere Project. (2011)** *Humanitarian charter and minimum standards in humanitarian response*. Geneva: Sphere Project. ([www.alnap.org/resource/8161.aspx](http://www.alnap.org/resource/8161.aspx)).

**Start Fund. (2014)** *South Sudan crisis response summary dashboard crisis response*. London: Start Fund. ([www.alnap.org/resource/19411.aspx](http://www.alnap.org/resource/19411.aspx)).

**Start, D. and Hovland, I. (2004)** *Tools for policy impact: A handbook for researchers*. London: ODI. ([www.alnap.org/resource/8479.aspx](http://www.alnap.org/resource/8479.aspx)).

**Steets, J., Darcy, J., Weingartner, L. and Leguene, P. (2014)** *Strategic evaluation FAO/WFP joint evaluation of food security cluster coordination in humanitarian action*. Rome: WFP. ([www.alnap.org/resource/19699.aspx](http://www.alnap.org/resource/19699.aspx)).

**Steets, J. and Dubai, K. (2010)** *Real-time evaluation of UNICEF's response to the Sa'ada conflict in Northern Yemen*. Berlin: GPPI. ([www.alnap.org/resource/8896.aspx](http://www.alnap.org/resource/8896.aspx)).

**Steets, J., Grünewald, F., Binder, A., de Geoffroy, V., Kauffmann, D., Krüger, S., Meier, C., and Sokpoh, B. (2010)** *Cluster approach evaluation 2 Synthesis Report*. Berlin: GPPi. ([www.alnap.org/resource/5986.aspx](http://www.alnap.org/resource/5986.aspx)).

**Stein, D. and Valters, C. (2012)** *Understanding theory of change in international development: a review of existing knowledge*. London: London School of Economics. ([www.alnap.org/resource/23088.aspx](http://www.alnap.org/resource/23088.aspx)).

**Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R., and Befani, B. (2012)** *Broadening the range of designs and methods for impact evaluations: Report of a study commissioned by the Department for International Development*. London: DFID. ([www.alnap.org/resource/8196.aspx](http://www.alnap.org/resource/8196.aspx)).

**Stetson, V. (2008)** *Communicating and reporting on an evaluation. Guidelines and tools*. Washington, DC: American Red Cross. ([www.alnap.org/resource/9888.aspx](http://www.alnap.org/resource/9888.aspx)).

**Stoddard, A., Harmer, A., Haver, K., Salomons, D. and Wheeler, V. (2007)** *Cluster approach evaluation*. New York: OCHA. ([www.alnap.org/resource/3641.aspx](http://www.alnap.org/resource/3641.aspx)).

**Stokke, K. (2007)** *Humanitarian response to natural disasters: a synthesis of evaluation findings*. Oslo: NORAD. ([www.alnap.org/resource/3551.aspx](http://www.alnap.org/resource/3551.aspx)).

**Street, A. (2009)** *Synthesis report: Review of the engagement of NGOs with the humanitarian reform process*. ([www.alnap.org/resource/12647.aspx](http://www.alnap.org/resource/12647.aspx)).

**Suchman, L. and Jordan, B. (1990)** 'Interactional troubles in face-to-face survey interviews'. *Journal of the American Statistical Association*, 85(409): 232–241. ([www.alnap.org/resource/23089.aspx](http://www.alnap.org/resource/23089.aspx)).

**Sutter, P., Downen, J., Walters, T., Izabiriza, B., Kagendo, R. S., Langworthy, M., Sagara, B. and Mueller, M. (2012)** *The contribution of food assistance to Durable Solutions in protracted refugee situations: its impact and role in Rwanda (2007–2011)*. Rome: WFP. ([www.alnap.org/resource/19832.aspx](http://www.alnap.org/resource/19832.aspx)).

**TANGO International. (2007)** *End of project study of tsunami impacted communities in Southern India*. Tucson, AZ: TANGO International. ([www.alnap.org/resource/3536.aspx](http://www.alnap.org/resource/3536.aspx)).

**Taplin, D. H. and Heléne, C. (2012)** *Theory of change basics: A primer on theory of change*. New York: ActKnowledge. ([www.alnap.org/resource/22902.aspx](http://www.alnap.org/resource/22902.aspx)).

**Taplin, D. H., Clark, H., Collins, E. and Colby, D. C. (2013)** *Theory of change: Technical papers: a series of papers to support development of theories of change based on practice in the field*. New York: ActKnowledge. ([www.alnap.org/resource/23090.aspx](http://www.alnap.org/resource/23090.aspx)).

**Taylor-Powell, E. and Henert, E. (2008)** *Developing a logic model: Teaching and training guide*. Madison, WI: University of Wisconsin. ([www.alnap.org/resource/23091.aspx](http://www.alnap.org/resource/23091.aspx)).

**Tearfund. (2015)** *Impact and learning report 2015: Inspiring change*. Teddington: Tearfund. ([www.alnap.org/resource/23092.aspx](http://www.alnap.org/resource/23092.aspx)).

**Teddlie, C. and Yu, F. (2007)** 'Mixed methods sampling: A typology with examples'. *Journal of Mixed Methods Research*, 1(1): 77–100. ([www.alnap.org/resource/23093.aspx](http://www.alnap.org/resource/23093.aspx)).

**Telford, J. (1997)** *Counting and identification of beneficiary populations in emergencies: registration and its alternatives*. London: ODI. ([www.alnap.org/resource/11010.aspx](http://www.alnap.org/resource/11010.aspx)).

**Telford, J. (2009)** *Review of joint evaluations and the future of inter agency evaluations*. New York: OCHA. ([www.alnap.org/resource/5772.aspx](http://www.alnap.org/resource/5772.aspx)).

**Telford, J. and Cosgrave, J. (2007)** *The international humanitarian system and the 2004 Indian Ocean earthquake and tsunamis*. ([www.alnap.org/resource/8162.aspx](http://www.alnap.org/resource/8162.aspx)).

**Telford, J., Cosgrave, J. and Houghton, R. (2006)** *Joint evaluation of the international response to the Indian Ocean tsunami: Synthesis report*. London: TEC. ([www.alnap.org/resource/3535.aspx](http://www.alnap.org/resource/3535.aspx)).

**Telyukov, A. and Paterson, M. (2008)** *Impact evaluation of PRM humanitarian assistance to the repatriation and reintegration of Burundi refugees (2003-08)*. Washington, DC: USAID. ([www.alnap.org/resource/5661.aspx](http://www.alnap.org/resource/5661.aspx)).

**Ternström, B. (2013)** *Evaluation of five humanitarian programmes of the Norwegian Refugee Council and of the standby roster: NORCAP*. Oslo: NORAD. ([www.alnap.org/resource/10101.aspx](http://www.alnap.org/resource/10101.aspx)).

**Tessitore, S. (2013)** "Like a good trip to town without selling your animals": *A study of FAO Somalia's cash for work programme*. Rome: FAO. ([www.alnap.org/resource/8821.aspx](http://www.alnap.org/resource/8821.aspx)).

**Theis, J. and Grady, H. (1991)** *Participatory rapid appraisal for community development: A training manual based on experiences in the Middle East and North Africa*. London: IIED. ([www.alnap.org/resource/8163.aspx](http://www.alnap.org/resource/8163.aspx)).

- Thizy, D. (2013)** *Federation-wide livelihood program evaluation*. Geneva: IFRC. ([www.alnap.org/resource/8864.aspx](http://www.alnap.org/resource/8864.aspx)).
- Thomas, A. and Mohnan, G. (2007)** *Research skills for policy and development: how to find out fast*. Thousand Oaks, CA: SAGE. ([www.alnap.org/resource/8164.aspx](http://www.alnap.org/resource/8164.aspx)).
- Thomas, V. and Beck, T. (2010)** *Changing the way UNHCR does business? An evaluation of the age, gender and diversity mainstreaming strategy, 2004-2009*. Geneva: UNHCR. ([www.alnap.org/resource/5852.aspx](http://www.alnap.org/resource/5852.aspx)).
- Todd, D., Batchelor, C., Brouder, S., Coccia, F., Castiel, E. F. and Soussan, J. (2015)**  
Evaluation of the CGIAR research program on water, land and ecosystems. Montpellier: CGIAR. ([www.alnap.org/resource/22803.aspx](http://www.alnap.org/resource/22803.aspx)).
- Tolani-Brown, N., Mortensen, J. and Davis, J. (2010)** *Evaluation of the UNICEF education programme in Timor Leste 2003 – 2009*. New York: UNICEF. ([www.alnap.org/resource/6383.aspx](http://www.alnap.org/resource/6383.aspx)).
- Tomlinson, M., Solomon, W., Singh, Y., Doherty, T., Chopra, M., Ijumba, P., Tsai, A. and Jackson, D. (2009)** *The use of mobile phones as a data collection tool: A report from a household survey in South Africa*. London: BioMed Central. ([www.alnap.org/resource/8165.aspx](http://www.alnap.org/resource/8165.aspx)).
- Beck, T. (2011)** *Joint humanitarian impact evaluation: Report for the inter-agency working group on joint humanitarian impact evaluation*. London: TEC. ([www.alnap.org/resource/6046.aspx](http://www.alnap.org/resource/6046.aspx)).
- tools4dev. (2014)** *How to pretest and pilot a survey questionnaire*. ([www.alnap.org/resource/22804.aspx](http://www.alnap.org/resource/22804.aspx)).
- Tufte, E. R. (1990)** *Envisioning information*. Graphics Press. ([www.alnap.org/resource/8166.aspx](http://www.alnap.org/resource/8166.aspx)).
- Tufte, E. R. (1997)** *Visual explanations: images and quantities, evidence and narrative*. Graphics Press. ([www.alnap.org/resource/22805.aspx](http://www.alnap.org/resource/22805.aspx)).
- Tufte, E. R. and Graves-Morris, P. R. (2001)** *The visual display of quantitative information*. Graphics Press. ([www.alnap.org/resource/8167.aspx](http://www.alnap.org/resource/8167.aspx)).
- Tulane University. (2012)** *Haiti humanitarian assistance evaluation from a resilience perspective*. New Orleans, LA: Tulane University. ([www.alnap.org/resource/6377.aspx](http://www.alnap.org/resource/6377.aspx)).
- Tuttle, A. H., Tohyama, S., Ramsay, T., Kimmelman, J., Schweinhardt, P, Bennett, G., Mogil, J. (2015)** 'Increasing placebo responses over time in U.S. clinical trials of neuropathic pain'. *Pain*, 156(12): 2616-2626. ([www.alnap.org/resource/22806.aspx](http://www.alnap.org/resource/22806.aspx)).

**Ubels, J., Acquaye-Baddoo, N. A. and Fowler, A. (2010)** *Capacity development in practice*. London: Earthscan. ([www.alnap.org/resource/22808.aspx](http://www.alnap.org/resource/22808.aspx)).

**UN General Assembly. (1989)** *Convention on the rights of the child*. New York: UN General Assembly. ([www.alnap.org/resource/22809.aspx](http://www.alnap.org/resource/22809.aspx)).

**UN General Assembly. (1951)** *Convention relating to the status of refugees*. New York: UN General Assembly. ([www.alnap.org/resource/22810.aspx](http://www.alnap.org/resource/22810.aspx)).

**UNDP. (2009)** *Handbook on planning, monitoring and evaluating for development results*. New York: UNDP. ([www.alnap.org/resource/22812.aspx](http://www.alnap.org/resource/22812.aspx)).

**UNDP. (2011)** *Updated guidance on evaluation in the handbook on planning, monitoring and evaluation for development results (2009)*. New York: UNDP. ([www.alnap.org/resource/22813.aspx](http://www.alnap.org/resource/22813.aspx)).

**UNEG. (2005a)** *Norms for evaluation in the UN system*. New York: UNEG. ([www.alnap.org/resource/8215.aspx](http://www.alnap.org/resource/8215.aspx)).

**UNEG. (2005b)** *Standards for evaluation in the UN system*. New York: UNEG. ([www.alnap.org/resource/19410.aspx](http://www.alnap.org/resource/19410.aspx)).

**UNEG. (2008a)** *Core competencies for evaluators of the UN System*, New York: UNEG. ([www.alnap.org/resource/22814.aspx](http://www.alnap.org/resource/22814.aspx)).

**UNEG. (2008b)** *Evaluability assessments of the programme country pilots: Delivering as one UN*. New York: UNEG. ([www.alnap.org/resource/22815.aspx](http://www.alnap.org/resource/22815.aspx)).

**UNEG. (2010a)** *Good practice guidelines for follow up to evaluations*. New York: UNEG. ([www.alnap.org/resource/8446.aspx](http://www.alnap.org/resource/8446.aspx)).

**UNEG. (2010b)** *UNEG quality checklist for evaluation reports*. New York: UNEG. ([www.alnap.org/resource/22816.aspx](http://www.alnap.org/resource/22816.aspx)).

**UNEG. (2011)** *Integrating human rights and gender equality in evaluation - Towards UNEG guidance*. New York: UNEG. ([www.alnap.org/resource/8171.aspx](http://www.alnap.org/resource/8171.aspx)).

**UNEG. (2013a)** *Resource pack on joint evaluations*. New York: UNEG. ([www.alnap.org/resource/12613.aspx](http://www.alnap.org/resource/12613.aspx)).

**UNEG. (2013b)** *UNEG handbook for conducting evaluations of normative work in the UN system*. New York: UNEG. ([www.alnap.org/resource/22817.aspx](http://www.alnap.org/resource/22817.aspx)).

**UNEG. (2014)** *Integrating human rights and gender equality in evaluation - Towards UNEG guidance*. New York: UNEG. ([www.alnap.org/resource/19283.aspx](http://www.alnap.org/resource/19283.aspx)).

**UNEG HEIG. (2016)** *Reflecting Humanitarian Principles in Evaluation*. New York: UNEG. (<http://www.alnap.org/resource/23385.aspx>).

**UNHCR. (1997)** *Commodity distribution*. Geneva: UNHCR. ([www.alnap.org/resource/22818.aspx](http://www.alnap.org/resource/22818.aspx)).

**UNHCR. (2007)** *Handbook for emergencies*. New York: UNHCR. ([www.alnap.org/resource/8172.aspx](http://www.alnap.org/resource/8172.aspx)).

**UNHCR and WFP. (2013)** *Synthesis report of the joint WFP and UNHCR impact evaluations on the contribution of food assistance to durable solutions in protracted refugee situations*. Geneva and Rome: UNHCR/WFP. ([www.alnap.org/resource/12260.aspx](http://www.alnap.org/resource/12260.aspx)).

**UNICEF. (1998)** *The state of the world's children*. New York: UNICEF. ([www.alnap.org/resource/19280.aspx](http://www.alnap.org/resource/19280.aspx)).

**UNICEF. (2002)** *Children participating in research, monitoring and evaluation - Ethics and your responsibilities as a manager*. New York: UNICEF. ([www.alnap.org/resource/8174.aspx](http://www.alnap.org/resource/8174.aspx)).

**UNICEF. (2005)** *Emergency field handbook*. New York: UNICEF. ([www.alnap.org/resource/8173.aspx](http://www.alnap.org/resource/8173.aspx)).

**UNICEF. (2010)** *Evaluation of the UNICEF education programme in Timor Leste 2003 – 2009*. New York: UNICEF. ([www.alnap.org/resource/6383.aspx](http://www.alnap.org/resource/6383.aspx)).

**UNICEF. (2013a)** *Evaluability assessment of the peacebuilding, education and advocacy programme*. New York: UNICEF. ([www.alnap.org/resource/12679.aspx](http://www.alnap.org/resource/12679.aspx)).

**UNICEF. (2013b)** *Evaluation of UNICEF programmes to protect children in emergencies: Pakistan country case study*. New York: UNICEF. ([www.alnap.org/resource/22820.aspx](http://www.alnap.org/resource/22820.aspx)).

**UNICEF. (2013c)** *Evaluation of UNICEF programmes to protect children in emergencies. Synthesis report*. New York: UNICEF. ([www.alnap.org/resource/12491.aspx](http://www.alnap.org/resource/12491.aspx)).

**UNICEF. (2013d)** *Global evaluation reports oversight system (GEROS)*. New York: UNICEF. ([www.alnap.org/resource/7948.aspx](http://www.alnap.org/resource/7948.aspx)).

**UNICEF. (2013e)** *Thematic synthesis report on evaluation of humanitarian action*. New York: UNICEF. ([www.alnap.org/resource/8442.aspx](http://www.alnap.org/resource/8442.aspx)).

**UNICEF. (2013f)** *Improving Child Nutrition: The achievable imperative for global progress*. New York: UNICEF. (<http://www.alnap.org/resource/23386.aspx>).

**UNICEF. (2014)** *Evaluation of UNICEF's cluster lead agency role in humanitarian action (CLARE)*. New York: UNICEF. ([www.alnap.org/resource/19327.aspx](http://www.alnap.org/resource/19327.aspx)).

**UNICEF. (2015)** *Final Report for the formative evaluation of the highscope curriculum reform programme (February to December 2014)*. Christ Church: UNICEF. ([www.alnap.org/resource/22819.aspx](http://www.alnap.org/resource/22819.aspx)).

**UNICEF and Save the Children. (2010)** *Management response to the review of the global education cluster co-leadership arrangement*. New York and London: UNICEF/Save the Children. ([www.alnap.org/resource/6354.aspx](http://www.alnap.org/resource/6354.aspx)).

**UNIFEM. (2011)** *UNIFEM strategic plan 2008-2011: Evaluability assessment*. New York: UNIFEM. ([www.alnap.org/resource/22821.aspx](http://www.alnap.org/resource/22821.aspx)).

**Universalia. (2013)** *GEROS – Global meta-evaluation report 2012: final report*, New York: UNICEF. ([www.alnap.org/resource/22822.aspx](http://www.alnap.org/resource/22822.aspx)).

**UNODC. (2012)** *Evaluability assessment template*. Vienna: UNODC. ([www.alnap.org/resource/22823.aspx](http://www.alnap.org/resource/22823.aspx)).

**USAID. (1996)** *Preparing an evaluation scope of work*. Washington: USAID. ([www.alnap.org/resource/22825.aspx](http://www.alnap.org/resource/22825.aspx)).

**USAID. (2010)** *Performance monitoring and evaluation tips using rapid appraisal methods*. Washington, DC: USAID. ([www.alnap.org/resource/19281.aspx](http://www.alnap.org/resource/19281.aspx)).

**USAID, 2012.** *How-to note: Preparing evaluation reports*, Washington, DC: USAID. ([www.alnap.org/resource/22824.aspx](http://www.alnap.org/resource/22824.aspx)).

**USAID. (2013)** *After-action review guidance*. Washington, DC: USAID. ([www.alnap.org/resource/19257.aspx](http://www.alnap.org/resource/19257.aspx)).

**USAID. (2015)** *USAID office of food for peace food security desk review for Katanga, North Kivu and South Kivu, Democratic Republic of Congo*. Washington, DC: USAID. ([www.alnap.org/resource/22826.aspx](http://www.alnap.org/resource/22826.aspx)).

**Valters, C. (2014)** *Theories of change in international development: Communication, learning, or accountability?* London: London School of Economics. ([www.alnap.org/resource/19265.aspx](http://www.alnap.org/resource/19265.aspx)).

**Van Bruaene, M., Dumélie, R., Kunze, M., Pankhurst, M. and Potter, J. (2010)** *Review concerning the establishment of a European voluntary humanitarian aid corps*. Brussels: Prolog Consult. ([www.alnap.org/resource/3633.aspx](http://www.alnap.org/resource/3633.aspx)).

**Varanoa, S. P., Schaferb, J. A., Cancinoc, J. M., Deckerd, S., H. and Greenee, J. R. (2010)** 'A tale of three cities: Crime and displacement after Hurricane Katrina'. *Journal of Criminal Justice*, 38(1): 42-50. ([www.alnap.org/resource/22827.aspx](http://www.alnap.org/resource/22827.aspx)).

**Wagner, A. K., Soumerai, S. B., Zhang, F. and Ross-Degnan, D. (2002)** 'Segmented regression analysis of interrupted time series studies in medication use research'. *Journal of Clinical Pharmacy and Therapeutics*, 2002(27): 299-309. ([www.alnap.org/resource/22830.aspx](http://www.alnap.org/resource/22830.aspx)).

**Walden, V. (2013)** *A quick guide to monitoring, evaluation, accountability, and learning in fragile contexts*. Oxford: Oxfam. ([www.alnap.org/resource/10593.aspx](http://www.alnap.org/resource/10593.aspx)).

**Warner, A. (2014)** *EHA, let's do some stretches! ALNAP EHA Practice Note*. Shared in ALNAP Humanitarian Evaluation Community of Practice in October.

**Warner, A. (2014)** *Evaluation of humanitarian action discussion series. Repeat after me: communicate, disseminate and support take-up!* London: ALNAP. ([www.alnap.org/resource/12920.aspx](http://www.alnap.org/resource/12920.aspx)).

**Webb, E. J., Campbell, D. T., Schwartz, R. D. and Sechrest, L. (1966)** *Unobtrusive measures: nonreactive research in the social sciences*. Thousand Oaks, CA: SAGE. ([www.alnap.org/resource/22831.aspx](http://www.alnap.org/resource/22831.aspx)).

**Webster, C. (1996)** 'Hispanic and Anglo interviewer and respondent ethnicity and gender: The impact on survey response quality. *Journal of Marketing Research*, 33(1): 62–72. ([www.alnap.org/resource/8176.aspx](http://www.alnap.org/resource/8176.aspx)).

**Weingärtner, L., Otto, R. and Hoerz, T. (2011)** *Die deutsche humanitäre Hilfe im Ausland. Band I: Hauptbericht*. Bonn: BMZ. ([www.alnap.org/resource/22832.aspx](http://www.alnap.org/resource/22832.aspx)).

**Weiss, C. H. (1997)** *Evaluation methods for studying programs and policies*. Upper Saddle River, NJ: Prentice Hall. ([www.alnap.org/resource/8540.aspx](http://www.alnap.org/resource/8540.aspx)).

**Wesonga, D. (2013)** *Accelerated primary education support (APES) project: Final evaluation report*. London: Concern Worldwide. ([www.alnap.org/resource/12232.aspx](http://www.alnap.org/resource/12232.aspx)).

**WFP. (2002)** *Emergency field operations pocketbook*. Rome: WFP. ([www.alnap.org/resource/8177.aspx](http://www.alnap.org/resource/8177.aspx)).

**WFP. (2010)** *Evaluation Matrix*. Rome: WFP. ([www.alnap.org/resource/22833.aspx](http://www.alnap.org/resource/22833.aspx)).

**WFP. (2012)** *Mixed-method impact evaluation - The contribution of food assistance to durable solutions in protracted refugee situations: its impact and role in Bangladesh*. Rome: WFP. ([www.alnap.org/resource/19830.aspx](http://www.alnap.org/resource/19830.aspx)).

**WFP. (2013)** *Synthesis of the series of joint UNHCR-WFP impact evaluations of food assistance to refugees in protracted situations*. Rome: WFP. ([www.alnap.org/resource/19903](http://www.alnap.org/resource/19903)).

**WFP. (2014)** *Impact evaluations I. guidance for process and content*. Rome: WFP. ([www.alnap.org/resource/22834.aspx](http://www.alnap.org/resource/22834.aspx)).

**WFP. (2015)** *Evaluation quality assurance system (EQAS) guidelines for operation evaluations*. Rome: WFP. ([www.alnap.org/resource/22835.aspx](http://www.alnap.org/resource/22835.aspx)).

- White, H. (2009a)** *Some reflections on current debates in impact evaluation*. New Delhi: 3ie. ([www.alnap.org/resource/8178.aspx](http://www.alnap.org/resource/8178.aspx)).
- White, H. (2009b)** *Theory-based impact evaluation: principles and practice*. New Delhi: 3ie. ([www.alnap.org/resource/8179.aspx](http://www.alnap.org/resource/8179.aspx)).
- White, H., 2014.** *The counterfactual. Howard White International Initiative for Impact Evaluation*. New Delhi: 3ie. ([www.alnap.org/resource/22836.aspx](http://www.alnap.org/resource/22836.aspx)).
- White, H. and Sabarwal, S. (2014a)** *Developing and selecting measures of child well-being*. Florence: UNICEF. ([www.alnap.org/resource/22837.aspx](http://www.alnap.org/resource/22837.aspx)).
- White, H. and Sabarwal, S. (2014b)** *Modelling*. Florence: UNICEF. ([www.alnap.org/resource/22838.aspx](http://www.alnap.org/resource/22838.aspx)).
- White, H. and Sabarwal, S. (2014c)** *Quasi-experimental design and methods*. Florence: UNICEF. ([www.alnap.org/resource/22839.aspx](http://www.alnap.org/resource/22839.aspx)).
- White, H., Sabarwal, S. and de Hoop, T. (2014)** *Randomized controlled trials (RCTs)*. Florence: UNICEF. ([www.alnap.org/resource/22840.aspx](http://www.alnap.org/resource/22840.aspx)).
- van der Wijk, J. and Kazan, L. (2010)** *Evaluation of the DEC-funded CAFOD health and WASH project in the DRC*. London: CAFOD. ([www.alnap.org/resource/5849.aspx](http://www.alnap.org/resource/5849.aspx)).
- Wilding, J., Swift, J. and Hartung, H. (2009)** *Mid term evaluation of DG ECHO's regional drought decision in the Greater Horn of Africa: March - May 2009*. Brussels: ECHO. ([www.alnap.org/resource/5919.aspx](http://www.alnap.org/resource/5919.aspx)).
- Wiles, P., Chan, L. and Horwood, C. (1999)** *Evaluation of Danish humanitarian assistance to Afghanistan 1992-98*. Copenhagen: DANIDA. ([www.alnap.org/resource/2827.aspx](http://www.alnap.org/resource/2827.aspx)).
- Willitts-King, B, Taylor, G. and Barber, K. (2012)** *Process Review of the Common Humanitarian Fund for Somalia*. New York: OCHA (<http://www.alnap.org/resource/6340>).
- Wilson, P. and Reilly, D. (2007)** *Joint evaluation of their responses to the Yogyakarta earthquake July 2007*. Atlanta, GA: CARE International. ([www.alnap.org/resource/3611.aspx](http://www.alnap.org/resource/3611.aspx)).
- W.K. Kellogg Foundation. (1998)** *W.K. Kellogg Foundation evaluation handbook*. Albuquerque, NM: ([www.alnap.org/resource/22841.aspx](http://www.alnap.org/resource/22841.aspx)).
- World Bank. (2004)** *Citizen report card surveys - A note on the concept and methodology*. Washington, DC: World Bank. ([www.alnap.org/resource/8182.aspx](http://www.alnap.org/resource/8182.aspx)).

**World Bank. (2009)** *Making smart policy: using impact evaluation for policy making*. Washington, DC: World Bank. ([www.alnap.org/resource/8434.aspx](http://www.alnap.org/resource/8434.aspx)).

**World Bank. (2007a)** 'Participatory tools for micro-level poverty and social impact analysis', in *Tools for institutional, political, and social analysis of policy reform*. Washington, DC: World Bank. ([www.alnap.org/resource/22842.aspx](http://www.alnap.org/resource/22842.aspx)).

**World Bank. (2007b)** *Sourcebook for Evaluating Global and Regional Partnership Programs: Indicative Principles and Standards*, Washington: The Independent Evaluation Group of the World Bank. ([www.alnap.org/resource/22843.aspx](http://www.alnap.org/resource/22843.aspx)).

**Wortel, E. (2009)** *Humanitarians and their moral stance in war: the underlying values*. Geneva: ICRC. ([www.alnap.org/resource/8183.aspx](http://www.alnap.org/resource/8183.aspx)).

**Yang, C. H. Xirasagar, S., Chung, H. C., Huang, Y. T. and Lin, H. C. (2005)** 'Suicide trends following the Taiwan earthquake of 1999: empirical evidence and policy implications'. *Acta Psychiatrica Scandinavica*, 112: 442–448. ([www.alnap.org/resource/8184.aspx](http://www.alnap.org/resource/8184.aspx)).

**Yarborough, D., Shulha, L., Hopson, R. and Caruthers, F.** *The program evaluation standards*. Thousand Oaks, CA: SAGE. ([www.alnap.org/resource/22725.aspx](http://www.alnap.org/resource/22725.aspx)).

**Zaiiontz, C. (2015)** *Performing real statistical analysis using excel*. ([www.alnap.org/resource/22844.aspx](http://www.alnap.org/resource/22844.aspx)).