

# Research Evidence in the Humanitarian Sector

A PRACTICE GUIDE



## ACKNOWLEDGEMENTS

We are very grateful to those who participated in the Evidence Lounge in Geneva, Switzerland in September 2017 at which this guide was discussed, provided comments on earlier drafts of the Guide or did both: Prisca Benelli (Save the Children), Kate Bingley (Christian Aid), Jonathan Breckon (Alliance for Useful Evidence), Alyoscia D'Onofrio (International Rescue Committee - IRC), Tarah Friend (UK Department for International Development - DfID), Tara Karul (International Initiative for Impact Evaluation - 3ie), Paul Knox-Clarke (Active Learning Network for Accountability and Performance - ALNAP), Flurina Kuhn (UK Science & Innovation Network), David Loquercio (Core Humanitarian Standard on Quality and Accountability - CHS), Tim Martineau (Liverpool School of Tropical Medicine - LSTM), Virginia Murray (Public Health England - PHE), Alice Obrecht (ALNAP), Aninia Nadig (Sphere Project), Adina Rom (ETH Zurich), Rodolfo Rossi (International Committee of the Red Cross - ICRC), Eva Svoboda (Overseas Development Institute - ODI, and Humanitarian Policy Group - HPG), Pierluigi Testa (Terre des hommes), Erik Von Elm (Cochrane Switzerland), Karin Wendt (Humanitarian Exchange and Research Centre - HERE).

Furthermore, the following organisations contributed examples (including case studies) and other content: 3ie, ALNAP, Christian Aid, Evidence Aid, International Committee of the Red Cross, International Rescue Committee, Liverpool School of Tropical Medicine, London School of Hygiene and Tropical Medicine, Overseas Development Institute, Public Health England, Save the Children, and ShelterBox.



## CITATION

This document should be cited as:  
Blanchet K, Allen C, Breckon J, Davies P, Duclos D, Jansen J, Mthiyane H, Clarke M. (2018) Using Research Evidence in the Humanitarian Sector: A practice guide. London, UK: Evidence Aid, London School of Hygiene and Tropical Medicine and Nesta (Alliance for Useful Evidence).

## AUTHORS

This document was written by Karl Blanchet<sup>a</sup>, Claire Allen<sup>b</sup>, Jonathan Breckon<sup>c</sup>, Phil Davies<sup>b</sup>, Diane Duclos<sup>b</sup>, Jeroen Jansen<sup>b</sup>, Helen Mthiyane<sup>c</sup> and Mike Clarke<sup>b</sup> (**a** Health in Humanitarian Crisis Centre, London School of Hygiene and Tropical Medicine; **b** Evidence Aid; **c** Nesta, Alliance for Useful Evidence).

## FUNDING

Funding for this document was provided by the UK Science & Innovation Network (represented by SIN Switzerland, British Embassy Berne).

## CORRESPONDENCE

Correspondence about this document should be sent to Karl Blanchet, Director of the Health in Humanitarian Crises Centre, London School of Hygiene & Tropical Medicine, Tavistock Place, London, UK; and Mike Clarke, Research Director of Evidence Aid, Centre for Public Health, Queen's University Belfast, ICS Block A, Royal Hospitals, Belfast, UK.

Karl: [Karl.Blanchet@lshtm.ac.uk](mailto:Karl.Blanchet@lshtm.ac.uk)  
Mike: [mclarke@qub.ac.uk](mailto:mclarke@qub.ac.uk)

Cover Photo © Patrick Brown/UNICEF/Panos Pictures

## THE WORK ON THE GUIDE WAS CO-ORDINATED BY THREE ORGANISATIONS:

**Evidence Aid** ([www.evidenceaid.org](http://www.evidenceaid.org)) champions an evidence-based approach in the humanitarian sector. Its mission is to alleviate suffering and save lives by providing the best available evidence on the effectiveness of humanitarian action and enabling its use. Evidence Aid has developed a range of resources which are available free of charge from its website. These include systematic reviews of interventions for disasters, humanitarian crises and other major emergencies. Evidence Aid organises events to promote the use of evidence in the humanitarian sector, including a yearly Humanitarian Evidence Week in November in collaboration with the Centre for Evidence-Based Medicine at the University of Oxford, UK alongside Evidence Lounges. Evidence Lounges bring practitioners and members of the academic research community together to enhance collaborations and develop practical ways of using evidence in the humanitarian sector, including this guide.

**Health in Humanitarian Crisis Centre** (<http://crises.lshtm.ac.uk>) at the London School of Hygiene and Tropical Medicine generates primary research and training on public health in humanitarian crises, working closely with international humanitarian agencies and research centres in affected countries to address critical health challenges.

A four-year research and capacity-building programme, RECAP, was launched in 2018 and has been focusing on decision making and accountability in response to humanitarian crises and epidemics. The RECAP project is supported by UK Research and Innovation as part of the Global Challenges Research Fund, grant number ES/P010873/1. The Centre also offers courses for practitioners through modules on conflict and health, short courses and a free online MOOC on Health in Humanitarian Crises.

**Nesta** (Alliance for Useful Evidence) ([www.alliance4usefulevidence.org](http://www.alliance4usefulevidence.org)) is a global innovation foundation, based in the UK. It backs new ideas to tackle the big challenges of our time, from the pressures of an ageing population to stretched public services and a fast-changing jobs market. It has produced several guides relevant to the use of evidence to meet this challenge and this publication draws upon one of these: 'Using Research Evidence: A Practice Guide'. That guide was produced by Nesta's Innovation Skills Team and the Alliance for Useful Evidence and the present guide for the humanitarian sector was developed in consultation with them and with grateful acknowledgment to Nesta for allowing us to use some of their content.



---

## CONTENTS

---

<b>INTRODUCTION</b>	<b>4</b>		
Who might use this guide?	5	Beyond experiments: new approaches to evaluation	32
How to use this guide	5	Was it worth it? The value of cost-benefit and cost- effectiveness analyses	32
<b>SECTION A: What is evidence- informed decision making, and why focus on research?</b>	<b>6</b>	Key messages from part 1 of Section C	34
What is 'evidence' and why do we focus on research?	7	<b>Part 2: How do you judge the quality of research?</b>	<b>35</b>
The challenge of relying solely on professional opinion	9	Peer-review: how does it help decision makers?	35
Key messages from Section A	10	Defining 'high-quality' research	35
		High-quality qualitative research	37
<b>SECTION B: When can evidence help you?</b>	<b>12</b>	Avoid cherry picking by using systematic reviews	37
Humanitarian Response Cycle	12	Meta-analysis	39
Developing and implementing a new intervention	14	Forest plots	39
Adopting ideas from other regions or sectors	18	Rapid evidence assessments	40
Creating a theory of change	19	The importance of repetition and corroboration	40
Key messages from Section B	19	Review of reviews: how to judge high-quality bodies of evidence	41
		Key messages from part 2 of Section C	42
<b>SECTION C: What evidence should you choose?</b>	<b>21</b>		
<b>Part 1: Different types of research methods, designs and approaches</b>	<b>23</b>	<b>SECTION D: Where should you look for evidence?</b>	<b>43</b>
Experimental research – why all the fuss?	28	Searching for research evidence	43
Experimental research designs	28	Sources of systematic reviews and other research for the humanitarian sector	44
Practical and ethical issues with randomised trials	30	Key messages from Section D	45
When randomisation is not possible or acceptable	30		
Quasi-experimental designs	30		
		<b>ENDNOTES</b>	
		References and endnotes	<b>46</b>

# Introduction

Evidence coming from research and evaluation can help you understand what works, where, why and for whom. It can also tell you what does not work, and help you avoid repeating the failures of others by learning from evaluations of unsuccessful humanitarian programmes. Evidence can also guide the design of the most effective ways to deliver specific interventions.

## **Evidence challenges what we might think is common sense, perceived or received knowledge.**

For instance, it was long believed that severe acute malnutrition required specialised treatment in inpatient facilities with therapeutic products, even though this model posed many challenges to effective treatment for both health systems and patients. In the mid-1990s, ready-to-use therapeutic food was developed. In 2000, initial pilot projects began to test the Community Management of Acute Malnutrition (CMAM) approach during humanitarian emergencies.<sup>1</sup> It was found to be so effective that it was endorsed by United Nations agencies in 2007,<sup>2</sup> and is now considered the standard of care for managing acute malnutrition in emergency and development contexts.

As financial and other resources are limited and often insufficient in the humanitarian sector, we cannot afford to waste such resources on policies and programmes that do not work. Interventions in any sector can have both positive and negative outcomes, and both intended and unintended consequences. Even in cases when

investing money to solve a problem is supposed to do some good, we need to ask ourselves if that money could be spent more effectively elsewhere, for example on a different humanitarian intervention. Research evidence can help you make informed choices on the most effective interventions to deliver in an emergency context.

Evidence is a contested field, with differing opinions on what should be most valued or deemed most relevant to decision makers.<sup>3</sup> However, for the purpose of this practice guide, we emphasise research that is underpinned by scientific notions of proof, validity, reliability, and has minimised bias. Such research has the advantage of rigour, relevance and independence. We focus on population studies that aim to generate average answers, which then need to be contextualised, for example taking into account cultural factors (such as the use of male doctors to examine and treat female patients), and issues of feasibility (such as the ability to deliver an intervention in the aftermath of a disaster when access to populations or resources might be especially restricted).

We have created this guide to help you make best use of research evidence when you are in a humanitarian emergency or when you are planning for the next emergency. Our intention is to help you find and use evidence on interventions, actions and strategies that might help you make informed choices and decisions. This guide is not about how to generate more research evidence. It is about using and understanding what evidence exists and recognising when good evidence is lacking. It should help you build your confidence in compiling, assimilating, distilling, and interpreting a strong evidence base of existing research, and think about how you might go on to evaluate your own projects and commission research or evaluation.

### WHO MIGHT USE THIS GUIDE?

This practice guide is primarily aimed at humanitarian decision makers and practitioners working in the field or in the headquarters of donor, international, national, or non-governmental organisations. It will help with decisions about the financing, supervision, delivery or evaluation of humanitarian interventions. It is not aimed at trained evaluators and researchers, but instead seeks to foster demand for research evidence from wider audiences in the humanitarian sector.

### HOW TO USE THIS GUIDE

The guide is divided into four main sections:

#### SECTION A

##### What is evidence-informed decision making, and why focus on research?

This section discusses what we mean by evidence-informed decision making, and why research is an essential element of it.

#### SECTION B

##### When can evidence help you?

This section explores different scenarios in which using evidence can help you, as well as the types of evidence you might need at different stages of developing or implementing a new intervention or policy.

#### SECTION C

##### What evidence should you choose?

This section looks at different types of evidence and examines how to choose the most appropriate for your case. It also discusses how to judge the quality of evidence.

#### SECTION D

##### Where should you look for evidence?

This section offers advice and resources to help you find the right evidence to support your case.

# What is evidence-informed decision making, and why focus on research?

This section discusses what we mean by evidence-informed decision making, and why research is an essential element of it.

**To begin, let us be clear about what we do not mean. We are not talking about making decisions and choices by slavishly following rigid research conclusions. Professional judgement and other sources of information – such as feedback from your stakeholders – will always be important. This practice guide is not about replacing professional judgement but increasing evidence use in humanitarian action.**

A good start in defining what we mean is borrowed from medicine. More than two decades ago, David Sackett and his colleagues proposed the following definition that has stood the test of time:

*“Evidence-based medicine is the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients. The practice of evidence-based medicine means integrating individual clinical expertise with the best available external clinical evidence from systematic research”.*<sup>4</sup>

This attempt to define evidence-based medicine was not the first,<sup>5</sup> but it has been influential and is just as relevant to the humanitarian sector as it is to other sectors. It stresses how research can complement professional judgement or other sources

of information and recognises the importance of evidence on issues such as feasibility, preference and culture.

In a field such as the humanitarian sector where more and better evidence is required,<sup>6</sup> we could assume that any model of good decision making should be wary of relying solely on professional judgement that is not supported by scientific evidence. Later in this section, you will read about how we can all be ‘predictably irrational’ and – consciously or unconsciously – make errors in important judgements. We explore how to mitigate these errors of judgement in subsequent sections. However, other decision making models have also stressed the importance of blending knowledge of evidence with judgement. The humanitarian sector is a sensitive area where we need to be aware of international and local politics and the dynamics between the various actors involved in the delivery of humanitarian aid. This will, sometimes, determine access to evidence and information, but also how humanitarian aid is delivered. However, the importance of evidence remains and, as noted in an ALNAP report in 2014, *“the failure to generate and use evidence in policy and response makes humanitarian action less effective, less ethical and less accountable”.*<sup>7</sup>

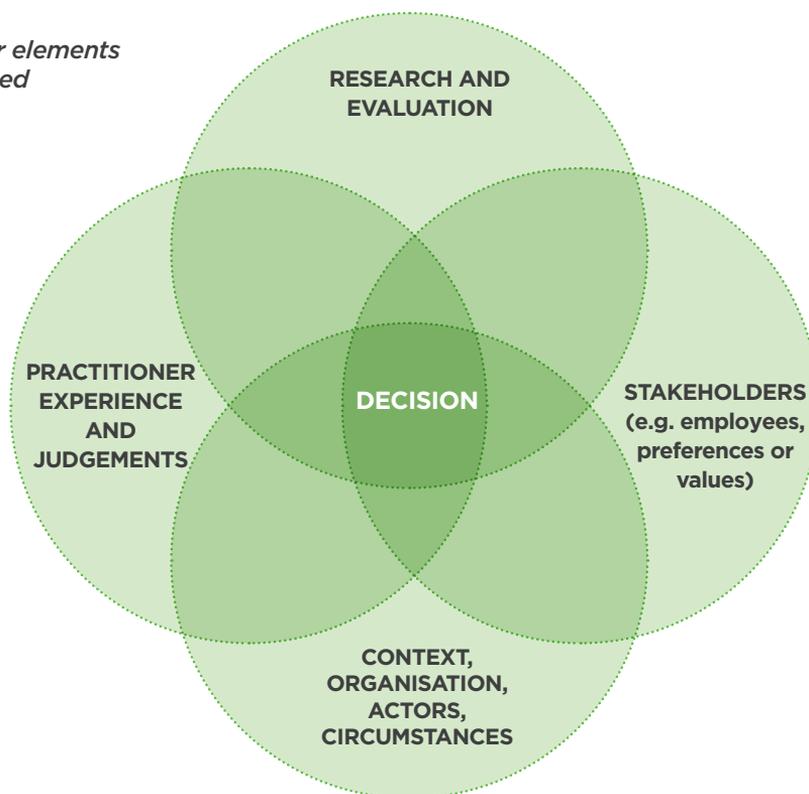
## WHAT IS 'EVIDENCE' AND WHY DO WE FOCUS ON RESEARCH?

The Oxford English Dictionary defines 'evidence' as *"the available body of facts or information indicating whether a belief or proposition is true or valid"*,<sup>8</sup> and, similarly, in their ALNAP report on the state of the evidence in the humanitarian sector, Paul Knox Clarke and James Darcy defined it as *"information that helps to substantiate or prove/disprove the truth of a specific proposition"*.<sup>7</sup> We follow these definitions because many other definitions tend to be rather unhelpful by being overly inclusive (sometimes including almost all types of information) or by being too abstract and vague.

Figure A.1 shows the different elements that should be part of evidence-informed decision making. Our focus in this practice guide is on the top circle of the diagram: research and evaluation.

As the authors of the Alliance for Useful Evidence's *'What Counts as Good Evidence?'* report state *"The conduct and publication of research involves the explicit documentation of methods, peer review and external scrutiny, resulting in rigour and openness. These features contribute to its systematic nature and help provide a means to judge the trustworthiness of findings. They also offer the potential to assess the validity of one claim compared to another"*.<sup>9</sup>

**Figure A.1:** The four elements of evidence-informed decision making



Based on: Barends E, Rousseau DM, Briner RB. (2014) Evidence-based Management: The Basic Principles. Amsterdam: Center for Evidence-Based Management [[www.cebma.org/wp-content/uploads/Evidence-Based-Practice-The-Basic-Principles.pdf](http://www.cebma.org/wp-content/uploads/Evidence-Based-Practice-The-Basic-Principles.pdf)]

This practice guide focuses on research, but there are many overlaps with the field of evaluation and we discuss some approaches to evaluating impact and process in Section C. We also give most attention to research that deals with impact – whether something has had positive or negative results – because questions on impact are vital to those involved in humanitarian action. These actors are concerned about showing their ‘impact’ on populations, their ‘results’ in international terms or ‘what works’ for governments and local and international providers. The language may change, but the idea for their research stays the same: to see if they have really made a difference. Therefore, our aim with this guide is to help you decide how that research might help you choose interventions, actions and strategies and adopt policies that are most likely to make a positive difference. We provide illustrative examples throughout the guide, and further examples of the use of evidence in the humanitarian sectors are available in other collections of case studies.<sup>10</sup>

We give prominence to research and evaluation that is ready-made, with no need to run a brand-new study. Decision makers have limited time and resources and many simply cannot afford to commission such a study and to wait for its results to become available, which may take years. Someone needing to make a decision now, needs the evidence now, if not yesterday, not in a year or more years’

time. So, decision makers require evidence that can be taken ‘off the shelf’ and combined with information on the local context to inform their choice. Fortunately, it is possible to find such evidence and we cover some of the ways to do so in Section D.

Research is a process engaged in for learning purposes. It seeks to answer questions such as ‘What was the commonest type of injury after an earthquake?’, ‘What are the effects on gender-based violence of different ways to protect women and children?’ or ‘How waterproof is a particular material when used for shelter?’

Evaluation is a process involving the assessment of findings and observations against standards, for the purpose of making decisions. Evaluations ask questions such as ‘Which types of first aid should first responders be trained in?’, ‘Which is the best way to protect women and children from gender-based violence?’ or ‘What material should be used for making tents in a setting with heavy rainfall?’

Research does not necessarily require evaluation. However, doing evaluation always requires doing research. An evaluation relates to an intervention that was actually implemented, while research is more comprehensive and, as well as including evaluations, it can also seek to answer conceptual questions, such as when planning for the needs that are likely after a disaster or developing

We give prominence to research and evaluation that is ready-made, with no need to run a brand-new study.

a novel intervention. However, whether we are thinking about research or evaluation, we need to be aware that bias can distort the findings. In Section C, we look at some of the biases that can affect the results of a study, but it is also important to be aware of the biases in how people interpret the results of a study or which influence their thinking about the need for research or its impact. We discuss some of these biases next.

### THE CHALLENGE OF RELYING SOLELY ON PROFESSIONAL OPINION

One reason we privilege good quality research over relying solely on professional judgement is that the biases involved in professional judgement can mean that it is wrong. As humans, we are ‘predictably irrational’<sup>11</sup> and may experience cognitive biases that distort our thinking. **Cognitive bias** refers to our inability to be entirely objective, which may manifest itself in multiples ways – such as perceptual distortion, inaccurate judgements and illogical or irrational interpretations.

Even with the best intentions, professionals can get it wrong. Take for instance one type of cognitive error: **confirmation bias**. This is the tendency to concentrate on the evidence that fits with what we believe, and to ignore or discount what does not. Even highly experienced professionals can fall into this cognitive trap.

Being aware of how we can jump to conclusions is important for making us wary of experts. But confirmation bias also highlights how anybody – not just experts – can be highly selective in their use of research evidence. People tend to look for the evidence that fits their beliefs, intentionally or not. This might be thought of as ‘policy-based evidence’ rather than ‘evidence-based policy’: cherry

picking and retro-fitting the evidence to the conclusions we have already fixed in our minds. If we are generous, this can be an unconscious mistake, but at worst it can be a deliberate ploy to back up prejudice by finding and promoting the evidence that fits that prejudice, while ignoring the stuff that does not.

A similar but different phenomenon is **optimism bias**, which refers to the “proven tendency for appraisers [of projects, programmes and policies] to be too optimistic about key project parameters, including capital costs, operating costs, project duration and benefits delivery”.<sup>12</sup> This can result in setting targets for outputs, outcomes and impacts that are unrealistic and undeliverable. Christoplos noted that, following the Indian Ocean Tsunami in 2004 “Methods and guidelines were filled with normative optimism. Terms such as the ‘seamless web’ were applied to describe the states of affairs that would emerge if the right methods were applied. Individuals, households, communities and nations were expected to follow a continuum from acute human suffering to ‘normal’ development. Over the years this optimism has faded. LRRD [Links between Relief, Rehabilitation and Development] has come to refer more to a conundrum than to a gilded path. Intractable political tensions and human insecurity have come to be associated with LRRD, first in post-conflict situations and then increasingly in natural disasters as well”.<sup>13</sup>

Optimism bias about both old and new interventions is often due to a lack of evidence about their true effects. Research and evidence from evaluations of these interventions, or similar ones, can help to reduce this uncertainty.

**Just as in other sectors, there are also many other biases relating to how people think that can afflict those working in the humanitarian sector. These include:**

---

**Hindsight bias:** Tendency to see past events as being more predictable than they were before the event occurred.

---

**Loss aversion:** Tendency to prefer avoiding losses than to acquiring gains.

---

**Framing effect:** Drawing different conclusions from the same information presented in different ways (e.g. would you prefer that '95% returned to work' or that '5% did not return to work'?).

---

**The 'availability heuristic':** When people relate the size, frequency or probability of a problem to how easy it is to remember or imagine.

---

**The 'representativeness heuristic':** When people overestimate the probability of vivid events.

---

**The 'need for coherence':** The urge to establish patterns and causal relationships when they may not exist.

---

**Meta-cognitive bias:** The belief that we are immune from biases!

This is not to say that professional judgement is always wrong. Researchers such as Gary Klein have sung the praises of intuitive expert judgement, for instance in his work on 'naturalistic decision making'.<sup>14</sup> Professional views and gut-instincts can be highly valuable, but we must be aware of their downsides. As Daniel Kahneman asserted in a joint article with Professor Klein in *American Psychologist*, "*professional intuition is sometimes marvellous, and sometimes flawed*".<sup>15</sup>

## CASE STUDY

 **How research has promoted the growth of  
Cash Transfer Programming**

In recent years, Cash Transfer Programming (CTP) has emerged as one of the most significant innovations in international humanitarian assistance.

In 2016, \$2.8 billion of humanitarian assistance was disbursed through cash and vouchers, a rise of 40% since 2015 and a doubling from 2014.<sup>16</sup> Cash disbursement has not only grown in total terms, but also as a relative share of total international humanitarian assistance, from 7.8% in 2015 to 10.3% in 2016.<sup>17</sup>

This expansion has been supported by a growing number of evaluations, resulting in a body of evidence on the effects of different programmes on individual and household-level outcomes. This body of research and evaluations has been instrumental in the growth of CTP.

It also helped to tackle concerns around cash transfers, such as corruption and insecurity,<sup>18</sup> and increased awareness that cash generates different, but not necessarily greater, risks than in-kind assistance.<sup>19</sup> The benefits of CTP have been shown to cut across multiple sectors to address women's empowerment, food security, education and health care. As a result, CTP is now an accepted tool in almost every emergency response. One of the key successes of CTP research is that the generated evidence has been widely used by champion institutions to promote the use of cash in humanitarian crises.

More research is nevertheless needed to fine-tune the programmes to maximise their benefits.

**Key messages from Section A**

✔ We do not advocate a form of decision making that slavishly follows the conclusions of research at the expense of professional judgement.

✔ However, you should not put professional expertise on a pedestal. Experts sometimes get it horribly wrong, and they are not immune from the whole range of social and cognitive biases that affect both experts and non-experts.

✔ Results from research are only one type of evidence, but have the advantages of greater rigour, relevance and independence when compared to other types of evidence.

✔ Creating new evidence can be costly and time-consuming, but there is good quality evidence that can be taken 'off the shelf' (see Section D). Research and evaluations of current emergencies can inform future interventions.

# When can evidence help you?

This section explores different scenarios in which using evidence can help you, as well as the types of evidence you might need at different stages of developing or implementing a new intervention or policy.

**Evidence can make organisations more effective. From more persuasive campaigning to securing grant funding and from developing an organisation's decision making abilities to making sure that programmes deliver results, evidence can bolster your work. It does not matter if you are a small voluntary organisation or a large international organisation. Whatever the scale, there may be existing research that can suit your needs.**

## EVIDENCE CAN HELP YOU IN MANY WAYS, INCLUDING:

- Starting doing something.
- Stopping doing something.
- Generating options or making the case.
- Identifying priorities.
- Assessing needs.
- Aligning services with needs.
- Designing and implementing more effective programmes.
- Developing funding bids.
- Creating effective advocacy campaigns.
- Increasing accountability to stakeholders.

## HUMANITARIAN RESPONSE CYCLE

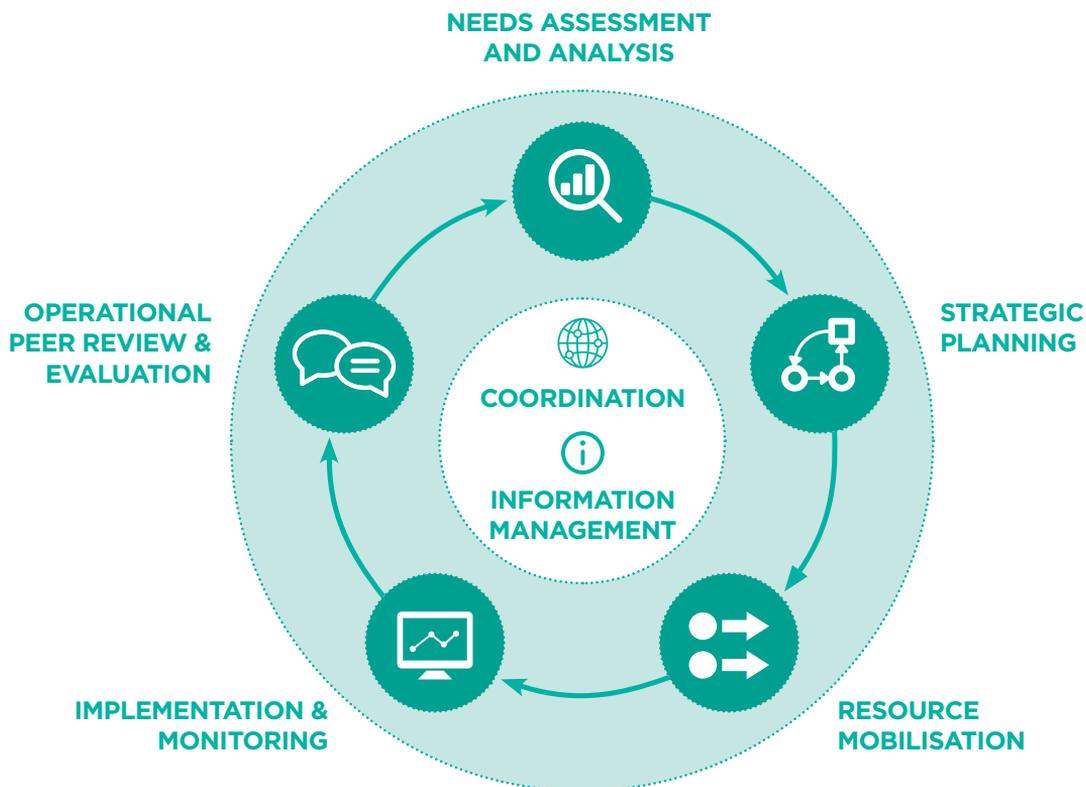
The United Nations Office for the Coordination of Humanitarian Affairs (UNOCHA) developed the humanitarian response cycle to show the coordinated series of actions needed to help prepare for, manage and deliver humanitarian action. It consists of five elements, with each step building on the previous and leading to the next.<sup>20</sup> Evidence can help with decision making as you move around this cycle.

### Needs assessment and analysis:

Looking at existing research will help provide a basis for assessing needs, and understanding the nature, magnitude and dynamics of specific problems. Along with the gathering of new data from the setting and circumstances you are confronted with (or expect to be confronted with), it can help to identify opportunities for responding appropriately. In some circumstances, this might be especially challenging if there have been few opportunities for relevant research.<sup>21</sup>

**Strategic response planning:** In planning your strategic response, you should search for evidence about what has been tried and assessed previously that might help you to make a well-informed decision. This might also include making use of evidence on how best to communicate the potential risks of a situation to those who might be involved.<sup>22</sup>

Figure B.1: Humanitarian Response Cycle



**Resource mobilisation:** The implementation of a new intervention or policy will require the mobilisation of resources. These might be financial to pay for activities or the personnel needed to deliver them. Having adopted an evidence-informed approach to planning the strategic response and choosing the interventions and policies to adopt, should make these more likely to be effective and, as a consequence, more attractive to donors.<sup>23</sup>

**Implementation and monitoring:** As you implement an intervention or policy you need to think about how you can monitor and report on your impact. This will help you to show if you are making a difference.

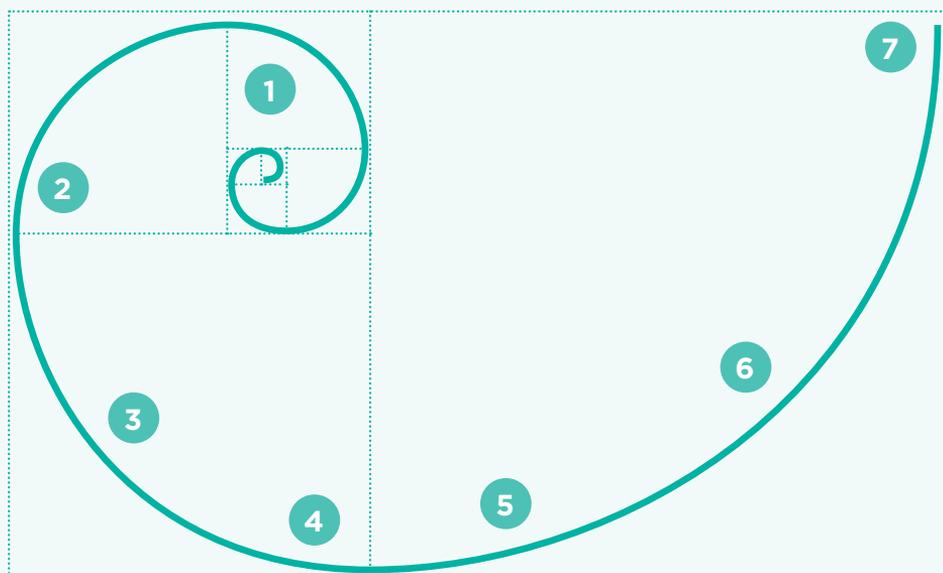
**Operational peer review and evaluation:** Operational peer review can be used as a course corrector and to identify areas for immediate corrective action. Along with an evaluation of your own strategic response, and comparisons with evidence from similar interventions, it can help to determine whether adjustments or improvements are necessary, in particular for leadership arrangements, implementation of other phases of the cycle, coordination and mechanisms for accountability to affected people.

## DEVELOPING AND IMPLEMENTING A NEW INTERVENTION

When developing or implementing a new humanitarian intervention or policy, it is a good idea to begin by thinking about timing. Different evidence will be helpful at different times in the development or implementation and you need to think about the appropriate research to suit different stages of the lifecycle of the new intervention or policy. In the early days of a new initiative, research can identify emerging challenges and the scale of the problem. It can be used to plot the development and implementation of a new intervention or policy.

The spiral below was developed by Nesta to capture the different stages of the innovation process and can also be used to plot the progress of a new approach to a social issue.<sup>24</sup> Different types of evidence will be needed at the different stages and the process should be continuous, such that when you have reached the end of the spiral, you should start again and being a new cycle. We discuss the various stages of this spiral below, while some specific examples of successful innovation in humanitarian contexts are available elsewhere from a study of 15 projects funded by the Humanitarian Innovation Fund.<sup>25</sup>

**Figure B.2: The Nesta Innovation Spiral**



- 1** Exploring opportunities and challenges
- 2** Generating ideas
- 3** Developing and testing

- 4** Making the case
- 5** Delivering and implementing
- 6** Growing and scaling
- 7** Changing systems

## 1. EXPLORING OPPORTUNITIES AND CHALLENGES

Working through current research will help you to understand the nature, magnitude and dynamics of the problems in front of you and the opportunities for responding appropriately. (See case study below.)

### CASE STUDY

#### Bringing evidence on long-term health systems outcomes to protracted crisis settings

With the principle of ‘leaving no-one behind’ embedded in the SDGs, there is agreement across most development partners that as well as addressing humanitarian health needs in crisis settings, this should be done in a way that supports longer-term equitable health system development.

The **ReBUILD Research Programme Consortium** is an international research partnership funded by UK Aid, working on improved access of the poor to effective health care in post-conflict settings through robust evidence on health systems strengthening. ReBUILD’s research has been conducted in relatively stable, if still fragile, post-conflict settings, using a historical lens to understand immediate, medium- and long-term impacts on the health system of policies implemented during and after conflict and crisis. Themes include health worker incentives and deployment, access to healthcare for the poorest, most vulnerable communities, aid effectiveness and gender equity.

The research has produced a rich understanding of the experience of communities and health workers through conflict and post-conflict periods, and recommendations for post-conflict health systems strengthening. But ReBUILD’s evidence is also relevant for humanitarian settings – e.g. the effects of approaches to engaging local health workers in emergency responses

on longer-term workforce distribution, motivation and skills, or the value of working with even weak local actors during emergencies, to support longer-term capacity for coordination, planning and predictable resource flows.

ReBUILD is therefore also engaging with those working on healthcare in current conflict or protracted crisis settings. Successes include informing decision-making and implementation in Sierra Leone during the Ebola outbreak, and supporting aspects of a donor’s support plan for Syria. Bringing evidence on long-term outcomes to those working in the humanitarian-development interface, where humanitarian needs are understandably the main focus, is challenging, with often limited space for long-term evidence to be considered within humanitarian organisations’ objectives and approaches, reflecting the wider divide between humanitarian and development approaches.

ReBUILD is collaborating with a number of organisations considering evidence-based humanitarian approaches, including Evidence Aid, and is now working closely with working groups with specific focus on the humanitarian-development interface for health, like the **UHC2030 Working Group on Fragile Settings** and the **Health Systems Global Thematic Working Group on Health Systems in Fragile and Conflict Affected States**.

## 2. GENERATING IDEAS

After you have identified your focus, it is time to search for interventions and policies that have been studied in the past. Can you borrow successful ideas from others about what worked previously and avoid interventions that failed?

## 3. DEVELOPING AND TESTING

As new ideas are initiated, it will be time to start thinking about testing and experimenting with different approaches and about evaluating the impact of these.

In Section C, we discuss Standards of Evidence that show the journey new programmes should go on in terms of providing evidence of impact.

## 4. MAKING THE CASE

Having evidence of the likely impact of your intervention will help you with making the case to those who might fund it and to those who will have to implement it in the field. This will put you in a stronger position to persuade them to fund the intervention and to support its delivery and implementation. (See case study below.)

### CASE STUDY

#### Presenting evidence to country offices leaders

In June 2017, Christine Fernandes, Save the Children’s global Humanitarian Nutrition Adviser for Infant and Young Children Feeding in Emergencies (IYCF-E) delivered a presentation about IYCF-E to Save the Children Iraq country director and senior leadership team, who were reviewing their 2017-2018 strategy. Up to that point, nutrition had not been one of the sectors of implementation in Iraq. Christine framed her 20-minute presentation as a quiz: each slide started with a question such as:

***Which do you think is the most effective means of preventing deaths in children under five?***

Insecticide treated materials.

Hib (meningitis) vaccine.

Appropriate breastfeeding.

Appropriate complementary feeding.

Vitamin A and Zinc supplementation.

She then proceeded to correct popular misperceptions with hard data. For example, by citing a Lancet article<sup>26</sup> to show that appropriate breastfeeding prevents the highest proportion of these deaths in early childhood (13%) and that complementary feeding has the second highest benefit. Asking people to make a guess, and then correcting their views is also an effective way for influencing people’s opinion, according to evidence.<sup>27</sup> Christine managed to persuade the whole Country Office leadership that supporting IYCF-E was the right thing to focus on, leading to a concrete change in the Iraq country office with an IYCF-E programme start-up in Baghdad and support to the Federal Ministry of Health on the national nutrition strategy focusing on breast milk substitute monitoring and capacity building on a national level in the early 2018 investment of resources.

## 5. DELIVERY AND IMPLEMENTATION

When you have implemented an intervention or policy you need to think about how you can document your impact. This will help take you from a position of rhetoric and saying ‘trust me, this project is working’ to one based on more trustworthy evidence from monitoring and evaluation. It will also help with your accountability by showing that you are making a difference, that your project is value for money, and that you have opened up your work to evaluation. (See case study below.)

### CASE STUDY

#### Partnering for impact evaluations

The International Initiative for Impact Evaluation (3ie) seeks to improve the lives of poor people in low- and middle-income countries by providing and summarising evidence of which development interventions work, when, why; and for how much. In 2018, 3ie’s Humanitarian Assistance Thematic Window is funding seven impact evaluations in Chad, Democratic Republic of the Congo (DRC), Mali, Niger, Pakistan, Sudan and Uganda. The impact evaluation in Pakistan is being conducted by the University of Mannheim In Germany, in collaboration with the Agency for Technical Cooperation and Development (ACTED). ACTED supports humanitarian assistance programmes in disaster-prone areas and the evaluation assesses the effectiveness of these interventions in responding to these hazards and reducing community vulnerability to emergency shocks. The collaboration between ACTED and the University of Mannheim has included several instances of successful capacity building and dissemination:

**Multi-layered engagement:** The study and baseline have been presented to the Pakistan government with the aim of informing Pakistan’s integrated nutrition strategy by drawing on

insights from various stakeholders with experience working on nutrition and disaster management. The findings have also been disseminated to an in-country working group comprising national and international NGOs, academia, and sector specialists.

**Contribution to improved data**

**quality:** There have been significant improvements in the Monitoring and Evaluation (M&E) system. Best practices on data collection such as spot checks, debriefing sessions with field teams and regular feedback on discrepancies in data collection, which were instituted during the impact evaluation by the research team have now been replicated in another project being implemented by ACTED.

ACTED hopes to include these practices in their regular M&E protocol when they have been further tested in the field.

**Building implementing agency**

**capacity on evaluation:** Staff at the implementing agency were sponsored to attend a workshop on measuring the effect of poverty reduction programmes and help shape better policies through the use of impact evaluations. This will enable them to better assimilate the findings from the evaluation, as well as provide a base for future in-house evaluations.

---

## 6) GROWING, SCALING AND SPREADING

---

The ultimate goal for many innovations is to replicate and grow so that they change things on a bigger scale in everyday practice and can be successfully copied in other settings. This will increase the scope for positive impact.

---

## 7) CHANGING SYSTEMS

---

Achieving system-wide changes is extremely complex. A good way to reach this scale is to have multiple independent studies showing that something does not just work in the specific circumstances of the early testing but is effective in many places.

## ADOPTING IDEAS FROM OTHER REGIONS OR SECTORS

Even if you find evidence of success of a policy or project elsewhere, would adoption of those ideas work in your area? Howard White gives an example of the failure of a supplementary feeding programme in Bangladesh (the Bangladesh Integrated Nutrition Project (BINP)) to achieve the outcomes of *“the acclaimed Tamil Nadu Integrated Nutrition Project (TINP) in India”*.<sup>28</sup> The Bangladesh programme had a 30% lower participation rate than that in India, and a counterfactual impact evaluation *“found no significant impact of the program on nutritional status, although there was a positive impact on the most malnourished children”*. Professor White noted that:

*“the people targeted have to be the right ones. The program targeted the mothers of young children. But mothers are frequently not the decision makers, and rarely the sole decision makers, with respect to the health and nutrition of their children. For a start, women do not go to market in rural Bangladesh; it is men who*

*do the shopping. And for women in joint households – meaning they live with their mother-in-law – as a sizeable minority do, then the mother-in-law heads the women’s domain. Indeed, project participation rates are significantly lower for women living with their mother-in-law in more conservative parts of the country.”*<sup>28</sup>

He also noted that the right children have to be identified for the programme to be effective. In the BINP areas, the community nutrition practitioners who implemented the programme *“could not correctly identify from the [growth] charts which children should be admitted to the program”*, which led to the mis-targeting of children.<sup>28</sup> These two examples of inappropriate programme implementation illustrate the importance of understanding local contextualized factors in the success or failure of a programme, and the need to use qualitative and ethnographic evidence (see Section C), in order to achieve desired outcomes in different contexts.

When designing new programmes aiming at changing how people act – for example, to adhere to International Humanitarian Law or to use latrines instead of defecating in open areas – research from social psychology, anthropology and behavioural change can help. It might provide ideas based on robust research about which of the range of techniques to ‘nudge’ people towards the desired actions would be worth adopting.

Hugh Waddington and colleagues,<sup>29</sup> for instance, noted that improving sanitation by getting people to use latrines rather than defecating openly in public spaces requires more than knowledge of germ theory. Evaluation evidence of Community Led Total Sanitation (CLTS) suggests that improved technical knowledge of latrines, the use of subsidies to reduce the costs to households and changing social norms and collective action at the village level are key factors in bringing about behavioural change and increasing latrine adoption.

## CREATING A THEORY OF CHANGE

In the early stages of developing or planning any intervention, it is important to logically describe what you do and why it matters. You need to be coherent, clear and convincing. This description is often referred to as a Theory of Change, and aims to give a “*clear, concise and convincing explanation of what you do, what impact you aim to have, and how you believe you will have it. It is a vital foundation of any programme, and a prerequisite for effective evaluation*”.<sup>30</sup>

A Theory of Change is a useful way to be more explicit about the evidence that you are using from others and to be clearer about how you are going to get results. For instance, if you want to run a programme to increase uptake of handwashing in a refugee camp, you may want evidence that answers a series of questions. What techniques are likely to lead to effective behavioural change? Will the new handwashing facilities be acceptable to potential users? Where should the

handwashing facilities be located? What should these facilities comprise of? What information needs to be provided? What incentives need to be introduced to influence household decisions?

A Theory of Change (example on page 20) helps you be explicit about your goals and how you will achieve those goals. It helps to avoid just simply hoping that your innovative approach will work and that your assumptions are correct. Another benefit of doing a Theory of Change is that it is a first step in designing an effective evaluation, because it tries to identify all the outcomes that will need to be measured.

The most important message here is to think about appropriateness. You want to find research that fits your needs and the stage of development that you are at. The evidence you will require in planning a new programme or during the first few months of its implementation will be very different to what you will need when the programme is more established. We revisit this crucial issue of appropriateness in Section C.

### Key messages from Section B

✔ There is a wide range of situations in which evidence can help you. This is not limited to just the obvious ones around capturing the results of interventions and policies. There may be other benefits that you have not thought about, such as creating more persuasive campaigning, obtaining funding or stopping doing something that is not working.

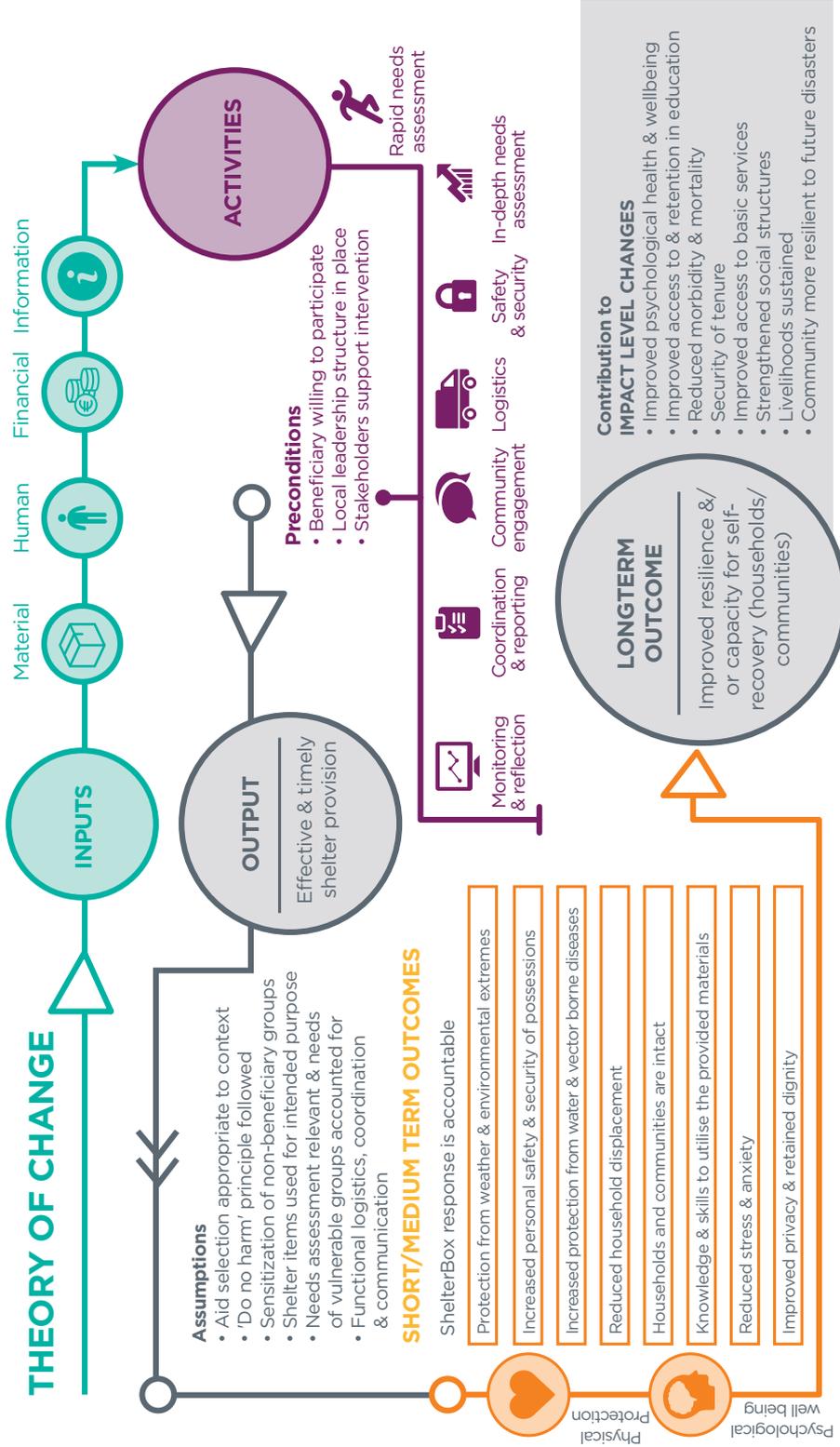
✔ You need to think about the timing and appropriateness of evidence. Where are you on the innovation spiral and what evidence may best suit your needs? For instance, in the early days of developing an innovative programme, it may be more valuable to learn from what others have already found, through examining past evaluations or wider social and scientific research, rather than doing a large, costly evaluation of your own work.

✔ Early stages of innovation are the time to create a Theory of Change. This is a useful way to be clearer about what evidence you are using and how you are going to get results. It will also help you to design an effective evaluation.

✔ Interventions or policies that are more established can set up their own experiments or evaluations of impact. As an innovation grows, you should also consider multiple replications to check that the benefits you saw were not just an isolated success story, but that it can work in other places and contexts.

✔ It does not matter what size of organisation you are in. There is a lot of research that might be able to help you.

Example of Theory of Change on resilience of communities



'Outcomes and impacts - measuring the difference we make': <https://cwarham.wordpress.com/2017/02/02/shelterbox-theory-of-change>



ShelterBox

# What evidence should you choose?

This section looks at different types of evidence and examines how to choose the most appropriate for the issue you are interested in. It also discusses how to judge the quality of evidence.

## **Not all evidence is equal.**

A key message of this practice guide is to think about usefulness. Too many people get into trouble by not thinking clearly about what sort of research design, methodology or methods of collecting data are really going to meet their needs. They tend to be 'methods led'.<sup>31</sup> In other words, we all have a tendency to pick our pet approach, such as questionnaires, randomised trials or interviews, because we are most comfortable with it, rather than being more open to the breadth of methods and thinking which ones are best suited to answer our challenge and resolve our uncertainties.

For instance, experimental research is more suited to evaluating impact and 'what works' (i.e. effectiveness), but other types of research can help give other insights.<sup>32</sup> They may not look so much at impact, but can reveal why and how things are working.<sup>33</sup>

Also, it might be important to establish which sections of the population benefit most from an intervention, and this can often be masked by relying on the average effect for the population as a whole.

CASE STUDY



## Lessons learnt from using ethnography in the Ebola response

The Ebola epidemic in West Africa shed light on the need to foster ‘community engagement’ when responding to outbreaks, and on the role played by networks of social scientists to share material and inform the response. This was exemplified by the Ebola Response Anthropology Platform, whose members *“work proactively with health and humanitarian organisations to design, deliver and monitor more locally responsive and socially informed interventions and research on the ground”*.<sup>34</sup> Lessons can be learnt from how anthropologists used ethnography to bring socio-cultural and political dimensions in a biomedical response in order to address the issue of so-called ‘unsafe’ burial practices. In a context where corpses transmit the Ebola virus, exceptional measures were adopted where only certain aid workers were authorised to bury the corpses of the persons who had died from Ebola. Gatherings of local communities attending funerals of loved ones who had died from Ebola were perceived by the Ebola outbreak

responders as a traditional resistance<sup>35</sup> to modern medicine, and as a threat to biosecurity. Involving anthropologists who had extensive experience living in the countries affected by the epidemic was key to identify acceptable and safe practices (such as showing a picture of the corpse to families to address rumours around bodies disappearing<sup>36</sup>), as well as to overcome tensions arising from burial practices through collaborative approaches. A key lesson that emerged was to identify rationales explaining communities’ needs to bury their dead as opposed to labelling these practices as essentially ‘cultural’.<sup>37</sup> The Ebola response showed *“the need for new global mechanisms to be established that can rapidly mobilise all experts who can bring relevant local contextual, medical, epidemiological, and political information on global health emergencies”*.<sup>38</sup>

The type of research you chose as the source of evidence to help in your decision making needs to fit the needs of the challenge that you face.<sup>39</sup>

Experimental research is more suited to evaluating impact and ‘what works’.

## PART 1: DIFFERENT TYPES OF RESEARCH METHODS, DESIGNS AND APPROACHES

### So how do you go about understanding which research method is appropriate for your case?

If you are not a researcher, it can be daunting to come face-to-face with the jargon and endless lists of different approaches.<sup>40</sup> These issues are captured in a *How to note* from the UK’s Department for International Development about assessing evidence<sup>41</sup> and are just as relevant to the humanitarian sector as they are to international development:

Some types of research (such as experimental and quasi-experimental designs) are better suited for identifying the presence of a causal relationship.

Other types of research are more appropriate for explaining such causal relationships – see *Beyond experiments: new approaches to evaluation* on page 32.

Some designs (such as observational studies) are more useful for understanding political, social and environmental contexts.

Some of the pros and cons of various research designs are summarised in **Table C.1**.<sup>42</sup>

**Table C.1** *Different designs, methods and approaches to research evidence – a brief overview*

TYPES OF RESEARCH AND EVALUATION	WHAT IS IT?	PROS	CONS
<b>Ethnography</b>	An account by someone with direct experience of, or affected by, a particular issue. The objective is to collect in-depth information on people’s experience and perceptions.	Powerful and immediate; may give vivid detailed insights into events concealed from much of the population.	Difficult to verify and may lead to inflation of prevalence. Emotive first-person stories may inhibit critical appraisal and individual anecdotes may not be representative.
<b>Case study</b>	Detailed and intensive examination of a single case or multiple cases such as a community, family, organisation, sites, event or individual.	Easy and practical; can be used by practitioners and non-specialists; good for interventions that have already happened; might identify adverse consequences from intervention; helps to describe innovations; generates hypotheses for other studies.	Bad at inferring that an intervention causes an outcome; small size means hard to generalise to national or population level.

TYPES OF RESEARCH AND EVALUATION	WHAT IS IT?	PROS	CONS
<b>Case control studies</b>	Compares a group who have, say, a health condition with a group of people who do not have it, and looks back in time to see how characteristics of the two groups differ.	Require fewer resources to carry out than intervention studies; useful when randomised trials or prospective cohort studies are not practical (e.g. studies of cot death); may generate hypotheses that can be tested in more robust studies.	Rare in social policy (see closely related 'case-based' evaluation below for approach more common in social policy), more frequent in epidemiology and health; provide less evidence for causal inference than a randomised trial; high risk of various biases (e.g. recall bias, selection bias and interviewer bias).
<b>Process evaluation</b>	An approach to evaluation that looks at what happens compared with existing theories of change or causal pathways identified during an evaluation. Can be associated with realist evaluation; qualitative comparative analysis (QCA); contribution analysis; process tracing.	Strong on explanation of causes; can be used in messier areas of social and complex interventions where there may be many causes and context is important. Process evaluation will investigate mechanisms of change, contextual factors and implementation evaluation.	Lack of agreed methods; opportunities for bias; weak on estimating quantities or extent of impact; relatively little evaluation of the techniques used in contrast to the large body of literature criticising randomised trials.
<b>Economic evaluation</b>	Economic evaluations determine whether an intervention is an efficient use of resources and how it compares to other interventions both in terms of their costs and consequences. In the humanitarian sector, the most common methods are cost/benefit analysis and cost-effectiveness analysis.	Economic evaluation provides a good sense of the level of resources needed and costs per benefit/effectiveness produced, which is a great indicator for most policy makers and can be used for advocacy.	Economic evaluation involves several assumptions and sometimes proxy measures of cost, which relies on judgements. The measure of effectiveness needs to be provided by a robust comparative trial.

TYPES OF RESEARCH AND EVALUATION	WHAT IS IT?	PROS	CONS
<b>Cross-sectional</b>	A representative sample of people surveyed at one point in time. Although surveys such as questionnaires and structured interviews are commonly used in cross-sectional design, they are not the only way. Other methods include content analysis or analysis of official statistics.	Quantitative data can be examined to detect patterns or association; relatively cheap and ethical; survey can be repeated at intervals, illustrating changing trends over time (see Cohort/Longitudinal studies below); may generate hypotheses that can be tested in more robust studies.	Establishes association at best, not causality; rarely possible to attribute any measured change to the intervention, or to understand what would have happened in the absence of the intervention because any change could have been due to broader issues such as economic conditions, weather, or media campaigns, rather than the intervention. Other disadvantages are risk of recall bias, social desirability bias, researcher bias; unequal group sizes and unequal distribution of confounders.
<b>Cohort/ Longitudinal studies</b>	The same sample of people surveyed over several points over time, sometimes from childhood to old age.	Best source of evidence on association between childhood experience and adult outcomes; can give powerful support for certain early interventions; can be used to evaluate programmes that are implemented following randomised trials.	Data often emerges too late for effective policy-making; study members may drop out over time; expensive approach when maintained over decades.

TYPES OF RESEARCH AND EVALUATION	WHAT IS IT?	PROS	CONS
<p><b>Quasi-experimental design</b></p>	<p>Compares a group who have, say, a health condition with a group of people who do not have it, and looks back in time to see how characteristics of the two groups differ.</p>	<p>Can provide reasonably strong evidence of the relationship between the intervention and the measured outcomes; powerful method of exploring the impact of an intervention when randomisation is impossible; can be applied to large communities as well as groups; no need for randomisation from the start (ex-ante), which avoids some of the challenges and difficulties of randomisation.</p>	<p>Inability to ensure equivalence of groups and to prevent change over time can result in less reliable findings; matching techniques tend to require a lot of data in both intervention and comparison groups which can be time-consuming and expensive to collect; a good understanding is required of the factors that need to be matched (without this, it remains possible that there are systematic differences between the groups that are not being controlled for); these designs require complex analytical work and specialist knowledge.</p>

TYPES OF RESEARCH AND EVALUATION	WHAT IS IT?	PROS	CONS
<p><b>Randomised trial</b></p>	<p>One group receives an intervention while another does not receive that intervention or receives a different intervention; groups are formed using a random process, usually with the same chance of being allocated to either group.</p>	<p>Offers the most robust, reliable findings, which give confidence that any measured difference between groups are due to the intervention; random allocation should prevent systematic differences between groups; greater confidence in the effect size and the relationship between the intervention and outcome; internationally recognised approach.</p>	<p>Poor on taking context into account (e.g. cultural, institutional, historical and economic settings); difficult to do at a national population level; when used at small pilot level, this might not be relevant to national/ population level (although this is a risk for all designs); can be hard to manipulate variables to experiment in social policy (e.g. class, race or where you live); mistakes in randomisation can invalidate results; can be hard to persuade decision makers of benefits of this design; potential political and ethical issues over randomisation (e.g. some groups randomly getting potential beneficial intervention, and not others); can take more management time and more time to set up than quasi-experiments.</p>
<p><b>Systematic reviews, meta-analysis</b></p>	<p>Aggregation of results from eligible studies, with the eligibility criteria defined in advance and methodologies reviewed</p>	<p>Best source of reassurance that an intervention works (or does not); meta-analysis pools statistical results; large reviews carry considerable statistical power; is replicable by other researchers; can be applied to any kind of data or study.</p>	<p>Requires a sufficient number of robust studies in a given area; methodology less well developed for synthesising qualitative data and 'grey' literature; might produce misleading results if selective reporting biases affect the research that is available for the review.</p>

## EXPERIMENTAL RESEARCH – WHY ALL THE FUSS?

Experimental research such as randomised trials have received a lot of attention – some might say too much. So why have they attracted so much interest?

Frequently, practitioners and policy makers want to know whether an intervention has had an impact. Has there been a causal link between their new programme and the ‘impact’, ‘effect’ and ‘result’ at the end? They want to be able to attribute the effects that they have seen to the policy or intervention that is being tested. Experimental designs, even though they are sometimes difficult to implement in practice, have a better chance of showing this cause and effect link.

An experimental study could simply test and compare two groups of people before and after the intervention is given to one of the groups. But if these groups are not formed by randomisation, there is a risk that the groups might differ because of selection bias and that differences in their outcomes would then be due to these differences in their characteristics and not to the effects of the intervention being tested. For example, somebody who asks for the intervention may be more predisposed to positive outcomes. Perhaps they are more motivated, healthier, confident, and thus more likely to do well regardless of any intervention. On the other hand, someone wanting the intervention may be the most vulnerable and most likely to do badly regardless of any intervention. These biases will distort the results of the experiment and may mean that it provides a false estimate for the effect of intervention. Furthermore, the biases might be unpredictable and if we do not know the direction of the bias, we cannot simply adjust the results of the experiment to try to take account of them.<sup>43</sup>

Instead, to mitigate this bias, we can use a process of randomisation: after the population for the study has been identified and they have agreed to join the study, the individuals to be compared with each other are allocated randomly to the study groups that will be compared, for example by drawing lots.<sup>44</sup>

## EXPERIMENTAL RESEARCH DESIGNS

Choosing the appropriate experimental design, and conducting the study properly establishes the causal link between an intervention and an outcome. When random allocation is used to assign the participants to the intervention or the comparator group, this provides the most robust means for determining whether any difference in outcomes between the groups is due to the intervention or chance, and chance is minimised by doing a large enough study. The random allocation minimises the possibility of systematic differences between the groups, regardless of whether these are due to known or unknown factors.

The introduction of a control group against which to compare the intervention minimises several biases that normally complicate the evaluation process. For example, if you introduce a cash transfer scheme to increase immunisation rates and the use of facility-based deliveries, how would you know whether those receiving the extra cash would not have used the services anyway? You need to compare them to a group of people who do not receive the extra cash.

As with any research design, the strength of a randomised trial also requires that the study is conducted properly. This includes ensuring that the randomisation of individuals or units (e.g. schools, hospitals, neighbourhoods etc.) is done properly (‘allocation fidelity’), and that only the experimental group has been exposed to

the intervention being tested ('treatment fidelity'). A well-conducted randomised trial also requires sufficient number of participants and statistical power to ensure that other factors ('confounders') that could have caused the outcome are balanced between the experimental and control groups. A low drop-out rate ('attrition') is also necessary to minimise the risk of bias. It is important, therefore, to check the reporting of a randomised trial to ensure that these criteria of validity have been met.

Further information on how to establish whether a randomised trial has been conducted and reported properly can be found in guidance on the conduct and reporting of trials.<sup>45</sup>

The International Initiative for Impact Evaluation (3ie) identified evaluations of humanitarian interventions in conflict and reconciliation situations that have used randomised trials and quasi-experimental designs (see below).<sup>46</sup>

## CASE STUDY

### Establishing the effectiveness of humanitarian interventions using experimental methods

Communicable diseases are of particular concern in conflict and disaster-affected populations that reside in camp settings. In the acute emergency phase, diarrhoeal diseases can account for more than 40% of deaths among camp residents. Diarrhoeal diseases result from a variety of factors but, in the emergency context, adequate sanitation and water supply (in terms of both quantity and quality) is the primary means for diarrhoea reduction. Clear limitations exist in current water treatment technologies, and few products are capable of treating turbid water.

Shannon Doocy and Gilbert Burnham described the findings of a 12-week effectiveness study of point-of-use water treatment with a flocculant-disinfectant among 400 households in camps for displaced populations in Monrovia, Liberia.<sup>47</sup>

The study compared diarrhoea rates among households with flocculant

disinfectant water treatment and improved water storage (intervention group) to households with only improved storage (control group). The allocation of households to the experimental and control groups was by randomisation.

In intervention households, point-of-use water treatment with the flocculant-disinfectant plus improved storage reduced diarrhoea incidence by 90% and prevalence by 83%, when compared with control households with improved water storage alone. Among the intervention, residual chlorine levels met or exceeded Sphere standards in 85% (95% CI: 83.1–86.8) of observations with a 95% compliance rate.

Drs Doocy and Burnham concluded that a point-of-use flocculant-disinfectant significantly reduced the incidence and prevalence of diarrhoeal disease among households residing in camp settings in the emergency context.

## PRACTICAL AND ETHICAL ISSUES WITH RANDOMISED TRIALS

Having a control group leads to a common criticism of randomised trials, which is that they are unethical because only some of the population are given access to an experimental intervention that might be considered superior. This 'rationing' of the intervention is justifiable if we do not know whether it will do more good than harm and we are doing the trial to find out. However, if there is already convincing evidence that the intervention will be beneficial, we might still need to do a randomised trial to find out how beneficial it is and for whom. If so, there are ways round the criticism and the ethical challenge. For example, in a waiting list control randomised trial, those allocated to the control group will be offered the intervention either at the end of the trial or when their personal follow-up has been completed. Another option is the 'stepped wedge' or 'pipeline' design,<sup>48</sup> in which the order in which participants receive the intervention is determined at random, allowing it to be gradually phased in over time so that all participants have been given access to it by the end of the trial. This design was used for the Progres/Oportunidades study, one of the largest randomised trials in international development, which tested conditional cash transfers for the rural poor in Mexico.<sup>49</sup> Stepped wedge trials can be highly practical in a world of austerity, where there are not enough resources to do a full roll-out of a new programme all in one go. They allow a gradual roll-out to be done while, at the same time, using a randomised trial to evaluate its effectiveness. However, stepped wedge trials are complex and can be difficult to implement, and if the intervention turns out to be ineffective or harmful, it will have been rolled out to everyone in the trial. In the case of a behavioural or educational intervention or the restructuring of service delivery, it might then be difficult or impossible to remove or 'switch off' the intervention after the trial.

## WHEN RANDOMISATION IS NOT POSSIBLE OR ACCEPTABLE

In some circumstances, it might not be possible or acceptable to randomly assign people to the groups to be compared. For example, it might be impossible to use a randomised trial to compare the relative effects of different ways to co-ordinate the response to a major emergency or the impact of a national policy intended to improve the social inclusion of refugees.<sup>50</sup> In such cases, researchers might use a quasi-experimental design to investigate the link between the interventions and the outcomes. Kayvan Bozorgmehr and Oliver Razum did this to compare personal spending on health care among asylum-seekers and refugees in Germany who had either restricted or regular access to health care.<sup>51</sup> Similarly, Rodolfo Rossi and colleagues used surveys before and after a vaccination campaign to investigate its effect on vaccination coverage in young children in a crisis-affected area of Lebanon in 2015.<sup>52</sup>

## QUASI-EXPERIMENTAL DESIGNS

There is a wide variety of quasi-experimental designs for comparing the effects of interventions and these are often used when randomised trials are not feasible. Such studies are similar to randomised trials to the extent that they can be used to compare the outcomes for individuals or groups who receive an intervention or programme and those who do not receive it. However, the receipt or non-receipt of the intervention or programme would not be based on random allocation. Instead, other types of comparative design, such as interrupted time series or controlled before-and-after studies, or analysis, such as regression discontinuity or propensity score matching, are used.<sup>53</sup> It should be noted that before-and-after designs without a comparison group cannot establish the impact of an intervention or programme because in the absence of a counterfactual

(i.e. a comparison with what would have happened without the intervention), any observed impact on outcomes may be due to factors other than the intervention. Quasi-experimental designs have a

greater risk of bias than well-conducted randomised trials, but they might still allow strong causal inferences to be made in circumstances where a randomised trial would not be possible or acceptable.<sup>54</sup>

CASE STUDY



**Establishing the effectiveness of humanitarian interventions using a quasi-experimental design. The Contribution of Food Assistance to Durable Solutions in Protracted Refugee Situations**

This evaluation was part of a series of impact evaluations jointly commissioned by the World Food Programme (WFP) and the Office of the United Nations High Commissioner for Refugees (UNHCR) in 2011 and 2012 to assess the role of food assistance in, and its contribution to, self-reliance and durable solutions for the refugee and the refugee-affected populations.<sup>55</sup>

The evaluation covered the protracted refugee situation in southeast Bangladesh, where approximately 30,000 Rohingya refugees have been assisted in two official camps for more than two decades. In addition to the registered refugees, approximately 45,000 unregistered Rohingya reside in makeshift sites and more than 150,000 reside in host communities in Cox’s Bazar district. To evaluate impact in the most methodologically rigorous manner for the context, a quasi-experimental design was used. The evaluation design focused primarily on three population groups:

1. Refugees living in two official refugee camps;
2. Unregistered Rohingya living in two unofficial sites, the official camps or host communities;
3. Host communities.

The key quantitative comparison was between registered refugees who received food assistance and

unregistered Rohingya who did not. The primary evaluation question was: What are the differential impacts of long-term food assistance on the different Rohingya refugee and refugee-affected populations in Bangladesh?

The evaluation concluded that along with other forms of external assistance, food assistance was a contributing factor in households’ choice of economic activity and adoption of specific coping strategies. Compared with their unregistered Rohingya counterparts, registered refugees engaged in significantly different economic activities, including higher-skilled and less risky employment for overall higher wage rates. They also had significantly better wealth status based on asset accumulation. Food assistance was an integral component of their livelihoods, used mainly for consumption and as collateral and a value transfer for loans and mortgages. The value transfer of all external assistance in the camps enabled refugees to work less and to rely on this external assistance in times of crisis.

Despite these differences, all refugee and unregistered Rohingya groups relied on economic activity to support their livelihoods. Unregistered Rohingya employed a wider range of coping mechanisms, both positive and negative, and were a significant part of the region’s labour market.

## BEYOND EXPERIMENTS: NEW APPROACHES TO EVALUATION

When thinking about impact, we also need to think beyond a sole dependence on experiments such as randomised trials. Although trials and quasi-experimental designs may work well when there is a simple intervention that can be tested, interventions are often not simple. NGOs working in civil society, for instance, rarely work alone or have the chance to manipulate a clear experimental ‘treatment’ for a randomised trial. Evaluators are looking at other ways of doing things, approaches that involve moving beyond a simple one-to-one causality and require careful thinking about your ‘contribution’.

In the complex world of humanitarian action, it is unlikely that your programme alone is the necessary or sufficient condition for success. It might be just one factor among many - part of a ‘causal package’. Programme success depends on what else is going on or has gone on around you. In the past, a simple causal explanation of ‘what works’ may have sufficed but nowadays it is more common for evaluation researchers to ask ‘did our intervention make a difference? What would have happened without it?’

There is a range of approaches that are becoming more popular with evaluators, such as theory-based<sup>56</sup> and case-based designs, complexity theory, realist evaluation,<sup>57</sup> contribution analysis,<sup>58</sup> process tracing<sup>59</sup> or qualitative comparative analysis.<sup>60</sup> A common feature of these approaches is a recognition of the complexity of causality requiring the identification of the causal mechanisms across a range of specific cases or events, as well as elements in the causal chain that vary. This involves *“identifying under what conditions, and in what combinations with other causes, the observed effect of an intervention is likely to be produced,*

*unlocked or triggered”*.<sup>61</sup> For instance, in contribution analysis, the research does not attempt to prove that one factor – such as a specific policy – ‘caused’ the desired outcome, but rather it sets out to explore the contribution a policy is making to observed results. It recognises that effects are produced by several causes at the same time, not just one.

These approaches are not all new.<sup>62</sup> Indeed, having a good theory has arguably always been at the heart of good science. Therefore, using a ‘theory-based’ approach is not novel, but these ways of evaluating impact have grown in popularity and they do help evaluators address multiple causality.

## WAS IT WORTH IT? THE VALUE OF COST-BENEFIT AND COST-EFFECTIVENESS ANALYSES

Having determined and measured the effects of an intervention, it is important for the humanitarian sector, as with any other sector, to know whether its programmes and interventions provide value for money. When budgets are constrained, we need to make difficult financial decisions about whether one thing is more valuable than another. Policy makers, commissioners of services, grant-making bodies and charitable funders alike are asking for more and better impact measures for every pound, dollar or euro they spend. Economic appraisal techniques help create such insights on value.

There is a range of techniques for doing economic appraisal. A cost effectiveness analysis (CEA) identifies which are the better ways to achieve an outcome. Whilst, cost benefit analysis (CBA) tells us if there are better uses for the available resources. Cost-utility analysis (CUA) calculates which intervention produces the greatest sense of subjective well-being for the people affected by it.<sup>12</sup>

Wherever feasible, economic appraisal tries to attribute monetary values to all the inputs and activities that go into producing the outputs which, in turn, lead to the desired outcomes and their longer term economic, social and environmental impacts. This constitutes the cost analysis. The same procedures are then used to attribute monetary values to all the outcomes and impacts that the policy is hoping to achieve. Where there are negative outcomes and impacts these are deducted from the monetary value of any positive outcomes/impacts that have been achieved. This constitutes the benefits analysis.

Monetary values are derived as much as possible from where there is some market activity. The labour market, for instance, provides monetary values for different types and units of labour. The building trade market provides the monetary value (the cost) for goods that are required to build a hospital or school – bricks, steel, glass, cement etc. The costs of providing medical care can be derived from the fixed and variable costs of running hospitals, health centres and community clinics, medicines and medical supplies, and the like. This is what is meant by ‘monetising’ costs and benefits.

Some costs and benefits (especially) are not traded in a market and are therefore more difficult to monetise. In such cases economists will estimate the benefits (and sometimes the costs) by simulating market activity with ‘willingness to pay’ and ‘willingness to accept’ surveys of people. Willingness to pay (WTP) has

been defined as “the maximum amount of money an individual is willing to give up in order to receive a good”.<sup>12</sup> Similarly, willingness to accept (WTA) “is the minimum amount of money they would need to be compensated to forego or give up a good”.<sup>12</sup> The results of a WTP or WTA survey provide what economists call ‘stated preferences’, that is, what people say they would be prepared to pay or accept for some good or service. Alternatively, economists will attempt to observe how people will actually behave in a real market situation. That is, they will set up a situation in which people are asked to pay different amounts of money for a good or service or will be offered different amounts of money to give up a good or service, and then record the actual behaviour of people. This provides what economists call ‘revealed preferences’.

In addition to the procedures mentioned above, economic appraisal also identifies where, or on whom, the costs and benefits will fall and over what time span. The economic value of benefits that will be realised in the medium to long-term future will be discounted by what is known at the discount rate. Other adjustments that are made to costs and benefits include changes in relative prices, material differences in tax options, optimism bias and contingencies such as floods, droughts and the effects of climate change. Economic appraisal is a fairly technical enterprise, usually requiring the expertise of an economist.

Quasi-experimental designs might still allow strong causal inferences to be made in circumstances where a randomised trial would not be possible or acceptable.

### Key messages for part 1 of Section C

- ✓ Not all evidence is equal. Some is better quality and will be more appropriate to your challenge.
- ✓ The type of research needs to fit the needs of the challenge and whichever method is used it needs to have been carried out and analysed appropriately.
- ✓ Do not be discouraged by the long list of research methods, designs and approaches. The key thing is to understand the assumptions that underpin these methods.
- ✓ Some research designs (such as experimental and quasi-experimental designs) are better suited for demonstrating the presence of a causal relationship.
- ✓ Other research approaches are more appropriate for explaining how such causal relationships come about or answering other types of research question.
- ✓ Theory-based evaluation and techniques such as contribution analysis are increasingly popular with evaluators. These techniques can be helpful when it is difficult to do an experiment or impossible to attribute your single policy or programme to any single clear result.
- ✓ It is not just about whether your intervention worked or not, but whether it was value for money.
- ✓ Techniques such as cost-benefit analysis can help you to understand the financial value of an intervention's impact.
- ✓ It might be important to establish which sections of the population benefit the most from an intervention. This can often be masked by relying on the average effect for the population as a whole.

It is important for the humanitarian sector to know whether its programmes and interventions provide value for money.

---

## PART 2: HOW DO YOU JUDGE THE QUALITY OF RESEARCH?

---

**Another way to help you choose which sort of research you need is to ask questions such as: what research can I trust? What is good enough evidence to fit my needs? Might it be an article in a high-impact scientific journal? What about the in-house evaluations conducted by my own organisation? Should they also have a place on the evidence table?**

Being published in a peer-reviewed research journal is one way to help you feel confident about a piece of research, but it is no guarantee of high quality. In a famous paper, John Ioannidis from Stanford University in the US caused a stir by arguing that ‘most published findings are probably false’. He examined the most cited papers (those with more than 1000 citations) in some of the best regarded medical journals in the world – largely drawn from *The Lancet*, *the New England Journal of Medicine* and *the Journal of the American Medical Association*.<sup>63</sup> Of those with claims of efficacy whose results had been tested in future studies, 41% were either found to be wrong, or the impact was much smaller than the original study had suggested.

### HOW DOES PEER REVIEW HELP DECISION MAKERS?

We must also be mindful that peer review, which is a cornerstone of academic journals is far from being perfect. There can be unconscious biases such as ‘herding’, where the behaviour of reviewers is influenced by the behaviour of their peers. And the whole system of scientific journals can be skewed by publication bias: positive results have a better chance of being published,<sup>64</sup> while negative data gets

hidden away. As an example, a survey by researchers at Stanford University found that most ‘null studies’ in the social sciences are never published: just 20% had appeared in a journal, and 65% had not even been written up.<sup>65</sup> This could have serious consequences for decision makers reviewing evidence. If you never see the negative studies on a new intervention, you may wrongly conclude that all is well and that it is effective.

Nevertheless, despite these problems, peer review remains, for most, the ‘gold standard’ for deciding what makes it into the scientific literature: providing a check against bad work.<sup>66</sup> Using a journal article or a report that has been independently peer-reviewed by other experts is one way of helping you be more confident that you can trust the research. Although it can take years before research gets published (and that can be too long if you have to make a quick decision), peer review brings some quality controls to the literature to help you feel confident.

### DEFINING ‘HIGH-QUALITY’ RESEARCH

Peer review may give us some modicum of comfort. But what do you do if you are going to include evidence that has not been checked by other experts? How do you decide if it provides a good enough basis for your decisions?

It would clearly be a mistake to ignore important research evidence just because it had not been published in a peer-reviewed journal. Or to miss the rich seam of donor or NGO evaluations that never get published in academic outlets but, instead, are available only in the ‘grey literature’. There are a wide variety of definitions for grey literature,

but it usually refers to documents that are unpublished or have been published without peer review. It can also refer to research that is still underway or being prepared for publication. Government reports, policy statements and briefs, and conference proceedings are also types of grey literature. Grey literature is important because it may contain evidence of negative outcomes and unsuccessful interventions, which is important for the balance of evidence-informed decision making. Grey literature can be searched using electronic databases such as Open Grey ([www.opengrey.eu](http://www.opengrey.eu)), conference proceedings and the procurement records of research funders. Websites of organisations that have an interest or expertise in a topic are another source of grey literature. These organisations can be contacted to identify researchers and decision makers who have expertise in a topic or issue.

A good start in trying to appraise the quality of evidence is defining it. One of the problems, however, is that phrases such as ‘quality’, ‘standards’, ‘robustness’, ‘bias’ and ‘strength’ are often used as if they were interchangeable, and without clearly defining what they mean. This makes for a lot of misunderstanding. For instance, in some guidance,<sup>67</sup> research ‘quality’ means using particular types of design and method – such as a randomised trial. This focus on minimising bias as a means of ensuring quality arises from some of the formal clinical and health approaches to assessing evidence

quality, such as the GRADE<sup>68</sup> or Maryland Scientific Methods Scale<sup>69</sup> systems. These approaches to quality assessment for experimental evaluations are usually based on the studies’ internal validity, quality of reporting and external validity.

Quality refers to how well studies have been conducted, reported and analysed,<sup>70</sup> as well as the researchers’ integrity in not distorting or falsifying their data.<sup>71</sup> Some people also link quality to how relevant the study is to policy and practice.<sup>72</sup>

When trying to answer a causal question, you need to consider whether the research design used for a study is appropriate for determining causality and whether the design was implemented properly in the study. High-quality impact evaluations will answer questions of attribution: showing that the intervention caused the outcomes. This requires a comparison or control group which is as similar as possible to the intervention group in all regards except the actual intervention. If this is true and the study has been well conducted, you can be more confident that, for example, the effects on the prevention of violence, reduction in family stress or faster return to work are due to the intervention. It is also important to consider whether the effects found in the study will be replicated in other places. This drives the demand for mixed methods of research and evaluation and might also require information from qualitative research.

It would clearly be a mistake to ignore important research evidence just because it had not been published in a peer-reviewed journal.

## HIGH-QUALITY QUALITATIVE RESEARCH

Many have highlighted the value of good quality qualitative research<sup>73</sup> for policy makers. Qualitative evidence can be just as scientifically credible as quantitative, and a combination of the two is likely to help decision makers combine evidence of the effects of an intervention with the knowledge to decide on the applicability of that evidence to their situation. The key thing is that research evidence that you use to deal with any challenge that you face is fit for purpose. It needs to have used the appropriate method, to have collected and analysed its data (whether qualitative or quantitative) using well-defined and replicable methods and to report the findings in a transparent way that minimises bias.

As with quantitative research, there are standards and guidelines for assessing the quality of qualitative research and evaluation.<sup>74</sup> Whereas the quality of experimental studies is usually based on internal validity, adequacy of reporting and external validity, the quality of qualitative research is usually built around the dimensions of contribution, defensibility, rigour and credibility. Contribution refers to whether the study advances wider knowledge or understanding about a policy, practice, theory or a particular substantive field. Defensibility refers to whether the study provides an appropriate research strategy to address the evaluative questions posed. Rigour refers to the systematic and transparent collection, analysis and interpretation of qualitative data. Credibility refers to how well-founded and plausible are the arguments about the evidence generated.

## AVOID 'CHERRY PICKING' BY USING SYSTEMATIC REVIEWS

Decision makers need to think about the quality of the whole body of evidence, not just single pieces of evidence. They need to use aggregated collections of

research and the accurate synthesis of existing information has been described as *“the most important single offering by academics to the policy process”*.<sup>75</sup> Much of the thinking on quality set out above has focused on single studies in ‘primary research’; in other words, individual studies such as experiments, surveys or a series of interviews. But, it’s important to think about ‘research syntheses’, and the summarising and pooling together of a series of individual studies. This idea is not new, and the concept of making decisions on the basis of accumulated evidence has been stressed for centuries.<sup>76</sup>

Our earlier discussion of cognitive biases showed how people can unconsciously fall into the trap of looking for evidence that fits our beliefs. So, we need to be careful when collecting together the existing research and need to be willing to include research even if we do not like its findings. One way to avoid ‘cherry picking’ is to use what are called systematic reviews.<sup>77</sup> These aim to be exhaustive, and to find as much as possible of the research relevant to answering the question that will help with the challenge we are faced with. They use explicit methods to identify what can reliably be said on the basis of the available studies and assess the research design and methods of these studies to determine their quality.

### Some of the key characteristics<sup>78</sup> of a systematic review are:

---

Clearly stated research question, objectives and eligibility criteria for studies.

---

Explicit and reproducible methodology to minimise bias.

---

Systematic search to identify as many studies as possible that meet the eligibility criteria.

---

Formal assessment of the validity of the findings of the included studies.

---

Systematic presentation, and synthesis, of the characteristics and findings of the included studies.

**CASE STUDY**

 **Using systematic reviews to develop guidance for staffs**

The International Rescue Committee (IRC) place a great reliance on evidence in their development of programme guidance documents and tools for field staff. They do this by conducting evidence reviews across many high-quality sources around specific interventions or approaches, and use what they learn about impact, contexts, populations and conditions to inform whether and how to adapt those interventions to specific contexts.

This is not without its challenges because most of the available studies do not provide critical information such as fidelity of implementation, causal mechanisms and gender sensitivity.

The IRC’s agency-wide effort to ensure that evidence is readily available to their staff is through the development of the Outcomes and Evidence Framework ([oef.rescue.org](http://oef.rescue.org)). In this electronic publicly available platform, the IRC have defined the outcomes and sub-outcomes that IRC focuses on in its work, the general theories of change (or pathways) through which they can achieve those outcomes and indicators for measuring them.

For each sub-outcome and outcome, the IRC have summarised the best available quantitative evidence on the effectiveness of interventions that aim to change/improve the relevant sub-outcome or outcome, with a primary focus on evidence from systematic reviews. For topics where systematic reviews do not yet exist, the IRC has identified and summarised individual impact evaluations.

Systematic approaches also have the value of being explicit about how they searched for research studies and decided on their eligibility. So, in theory at least,<sup>79</sup> other people could replicate the systematic review.

In 2018, Christl Donnelly and colleagues suggested four principles to help researchers, policymakers and others to commission, do, share, appraise and use evidence syntheses, including systematic reviews.<sup>80</sup> They proposed that evidence synthesis should be inclusive, rigorous, transparent and accessible and elaborated on these four areas as follows:

**Inclusive**

Involves policy makers and is relevant and useful to them.

Considers many types and sources of evidence.

Uses a range of skills and people.

Rigorous.

Uses the most comprehensive feasible body of evidence.

Recognises and minimises bias.

Is independently reviewed as part of a quality-assurance process.

**Transparent**

Clearly describes the research question, methods, sources of evidence and quality-assurance process.

Communicates complexities and areas of contention.

Acknowledges assumptions, limitations and uncertainties, including any evidence gaps.

Declares personal, political and organisational interests and manages any conflicts.

**Accessible**

Is written in plain language.

Is available in a suitable time frame.

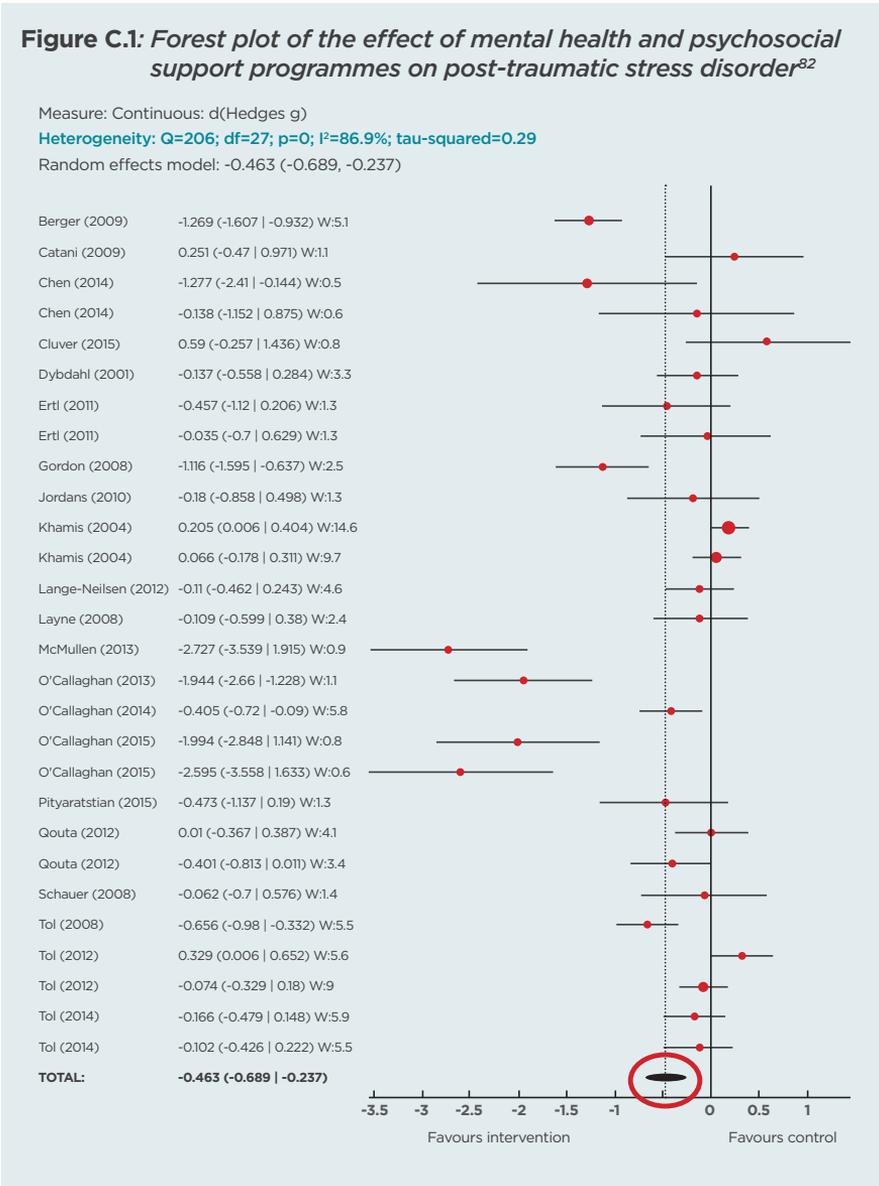
Is freely available online.

**META-ANALYSIS**

Meta-analysis is an important part of systematic reviews where all or some of the included studies are sufficiently similar ('homogeneous') in terms of population, intervention, comparators and outcomes for their statistical findings to be pooled and aggregated. This results in a cumulative estimate of effect which generally has a lower risk of bias than the effect estimate for individual studies in isolation and greater precision. Meta-analysis is usually represented by a forest plot<sup>81</sup> such as that in **Figure C.1**, which is taken from a recent systematic review on the impact of support programmes for populations affected by humanitarian emergencies. Meta-analysis is perhaps best known for combining the results of randomised trials, but it can also be done with data from other types of study, such as those that have used case-control, cross-sectional or cohort designs.

**FOREST PLOTS**

**Figure C.1** presents the findings of 28 impact evaluations that compared people in humanitarian emergencies who had received mental health and psychosocial support programmes with those who had not received this type of intervention.<sup>82</sup> In this forest plot, the data presents the effects of the programmes on the severity of post-traumatic



stress disorders (PTSD) as a continuous variable. For each study in this forest plot, the red dot represents the average treatment effect of the intervention, and the parallel lines either side of the red dots represent the confidence interval for that study. The solid black vertical line running from 0 on the horizontal axis indicates no difference between using and not using the programme, and the results of all the studies are pooled to provide the overall estimate of the effects of the programmes. This new, summary statistic is the black diamond (circled in red) at the bottom of the forest plot. This represents the *cumulative estimate of effect* of pooling and aggregating the average effects sizes and the variances of all 28 impact evaluations included in the review. It allows us to conclude that, on average, the mental health and psychosocial support programmes have a small, positive effect on PTSD compared with not using these interventions.

## RAPID EVIDENCE ASSESSMENTS

The preparation of systematic reviews and meta-analyses can be time-consuming. This means that if an up-to-date systematic review is not available, people needing to make an urgent decision might need to conduct their own searches for the relevant pieces of evidence and then appraise and synthesise this faster than would happen in a formal systematic review. Fortunately, the already large number of systematic reviews is continuing to grow rapidly, and we describe how to find them and several collations that are available in Section D. However, if you are unable to find what you are looking for amongst the existing systematic reviews, or the reviews you find are out of date, you might need to think about commissioning a 'pared-down systematic review', such as a rapid evidence assessment.<sup>83</sup> These rapid reviews normally take 1-3 months and are timed to meet the needs of policy makers and practitioners who cannot wait for a full

systematic review. They use the same basic structure and stages of full systematic review, but are not as intensive, exhaustive or comprehensive. They will take more short cuts with the searching, critical appraisal, data extraction and gathering, and statistical analysis of included studies. The findings are also presented in a shorter and less detailed form than a full systematic review, and might be no longer than 25 pages, with a three-page executive summary and a one-page briefing document for decision makers.

The limitations of rapid evidence assessments are that they are not as comprehensive or exhaustive as systematic reviews and are more likely to be subject to bias than a full systematic review. Consequently, greater caution is needed when basing a decision on evidence from a rapid evidence assessment than from a full systematic review. Notwithstanding these limitations, they are frequently commissioned and used by policy makers and programme implementers, especially where time is of the essence and no systematic reviews are available.

A particular type of rapid review, called a Rapid Research Needs Assessment, can also be used to quickly identify evidence gaps. The UK's Public Health Rapid Support Team for disease outbreaks includes a plan to conduct these assessments with Evidence Aid, to identify important uncertainties that could be tackled by research in the early stages of a humanitarian emergency associated with a disease outbreak.

## THE IMPORTANCE OF REPETITION AND CORROBORATION

Thinking about the concept of evidence gaps, brings us to one of the other things that needs to be considered when assessing the quality of a summary of research studies: the number of studies that need to be included for you to be comfortable that the body

of evidence is strong enough. Would only a couple of studies be sufficient if they are really good? Or might you require dozens, or even hundreds?

The reality is that there is no magic number of studies. Yet, we cannot ignore the fact that the size of the body of evidence is important: there is strength in numbers, and we must have repetition and corroboration. Even studies that have won many accolades need to be repeated. Amgen, a Californian drug company, tried to replicate 53 landmark cancer studies. The work, published in *Nature*,<sup>84</sup> only confirmed the findings of six (11%) of these important studies. This is deeply troubling because the studies have influenced drug companies and cancer treatments globally and the inability to replicate them might indicate that their findings are unreliable. On the other hand, we do not want to see studies being done again and again long after the evidence base is robust.<sup>85</sup> Unnecessary studies represent research waste<sup>86</sup> and, in the context of randomised trials, may be unethical if some participants continue to be randomised to sub-optimal interventions.

**A How to note on judging the strength of evidence produced by the UK's Department for International Development lists four things to consider when checking bodies of evidence:<sup>41</sup>**

The (technical) quality of the studies constituting the body of evidence (or the degree to which risk of bias has been addressed).

The size of the body of evidence.

The context in which the evidence is set.

The consistency of the findings produced by studies constituting the body of evidence.

## REVIEW OF REVIEWS: HOW TO JUDGE HIGH-QUALITY BODIES OF EVIDENCE

We need to remember that the formal ways of bringing research together, for example in systematic reviews, do not always lead to a high-quality answer. You still need to judge the quality of the design and implementation of the evidence review, and the quality of any synthesis can only be as good as the quality of the studies it is based on. Even the best-done review cannot turn low-quality research into a high-quality answer.

**There are also formal appraisal tools for assessing the quality of systematic reviews, such as AMSTAR.<sup>87</sup> You might also use one of the original checklists for assessing the quality of systematic reviews,<sup>78</sup> which posed a series of questions:**

Is the question clearly focused?

Is the search for relevant studies thorough?

Are the inclusion criteria appropriate?

Is the validity of the included studies adequately addressed?

Is missing information obtained from the original researchers?

How sensitive are the results to changes in the way the review is done?

Are subgroup analyses interpreted cautiously?

Do the conclusions flow from the evidence that is reviewed?

Are recommendations linked to the strength of the evidence?

Are judgements about preferences (values) valid?

Is 'evidence of no effect' confused with 'no evidence of effect'?

Judging all these criteria is always going to be rather subjective and needs to take account of the context of the policy question. It is difficult to give blanket advice for what would constitute the right body of evidence for any policy. We are also beginning to see the combination of the findings from multiple systematic reviews in 'reviews of reviews'. Of course, this cannot work if there is not a body of

existing systematic reviews – and in some sectors there are few. However, these reviews of reviews can provide an appraisal and summary of evidence from multiple systematic reviews of the same intervention; or provide an overview of the evidence from a collection of reviews of different interventions in the same topic area.<sup>88</sup>

### Key messages for part 2 of Section C

- ✔ To find evidence that you can trust, look for peer-reviewed research. But note that peer review is far from perfect. There can be unconscious biases such as 'herding', or publication bias towards positive results, or even the deliberate distortion or falsification of data.
- ✔ When looking at questions of impact and 'what works', use the frameworks and formal standards of evidence such as those used by Nesta and others.
- ✔ One study is never enough. Avoid making decisions based on single studies and look for multiple replications. There is strength in numbers, and we need repetition and corroboration. Even studies that have won many accolades need to be repeated, but we need to avoid doing studies again and again long after the evidence base is strong enough.
- ✔ Use systematic reviews, which aim to be exhaustive and screen studies for quality – usually based on the research design and methods.
- ✔ If you do not have time to do a systematic review or cannot find a completed one that answers your question, you might wish to conduct or commission a rapid evidence assessment.<sup>89</sup>

Even the best done review cannot turn low-quality research into a high-quality answer.

# Where should you look for evidence?

This section looks at different types of evidence and examines how to choose the most appropriate for the issue you are interested in. It also discusses how to judge the quality of evidence.

## SEARCHING FOR RESEARCH EVIDENCE

With the rapid growth in the number of people and organisations working on systematic reviews, the literature now contains a large and ever-increasing number of systematic reviews, so you might be able to find one that meets your needs. A recent estimate is that there are more than 200,000 systematic reviews across all topic areas. Of course, only a small proportion of these would be relevant to the humanitarian sector but (as noted below) bundles of such reviews are freely available in online collections such as those curated by Evidence Aid. There are also large international organisations, such as the Cochrane ([www.cochrane.org](http://www.cochrane.org)) and Campbell Collaborations ([www.campbellcollaboration.org](http://www.campbellcollaboration.org)) dedicated to the preparation and maintenance of systematic reviews who make these available online.

However, if you need to do your own searches for studies of the effects of humanitarian action, the starting point is to recognise that these are not always easy to find and you might need to get help from a librarian or information specialist who will be to advise and, perhaps, assist with designing and running the search. The evidence is scattered across tens of thousands of reports spread across thousands of

journals, books and websites; many of which are not free to use or are difficult to search. When thinking about where to search, it is important to consider where the evidence you are interested in is likely to have been published and indexed, and how comprehensive you wish to be. For example, if they were published in a scientific journal, the articles might be available through one of the thousands of electronic bibliographic databases. These include, for instance, PubMed for health care, LILACS for articles with particular relevance to South America or ERIC for educational literature. It may also be necessary to identify grey literature, such as government or NGO reports, and research presented at conferences, which will require searches of electronic databases such as Open Grey ([www.opengrey.eu](http://www.opengrey.eu)) and the websites of organisations that may have conducted relevant research.

In planning a search, it can be helpful to divide the search elements into the types of intervention or policy you are interested in, the relevant population or settings, the outcome measures that would be most helpful to your decision making and the types of study you wish to find. You can then decide which one or more of these domains is most important and helpful for identifying the material that is most relevant to you.

You should consider the types of source to search, which might include:

---

Bibliographic databases (e.g. PubMed, LILACS, ERIC).

---

Journal and conference websites (e.g. those with a particular focus on disasters or humanitarian action).

---

Registries of research (such as through the WHO portal for trial, [www.researchregistry.com](http://www.researchregistry.com) and PROSPERO for systematic reviews).

---

Online collections of research evidence (see below, e.g. ALNAP, 3ie).

---

Website of relevant organisations (e.g. for government and NGO reports).

---

References in articles.

---

Correspondence with researchers and evaluators.

#### When choosing the terms to include in your search, you should consider:

---

Synonyms from different times and places.

---

Other words and phrases that are related to what you're interested in.

---

Words that are broader.

---

Words that are more narrow or focused.

---

Index terms or keywords assigned to articles by the original authors or the bibliographic database.

---

And, finally, you need to decide on whether you will apply any restrictions based on language and the time period in which the research was conducted or published.

## SOURCES OF SYSTEMATIC REVIEWS AND OTHER RESEARCH FOR THE HUMANITARIAN SECTOR

Because of these challenges in seeking evidence, you might find it more efficient and easier to use trusted repositories of research and systematic reviews, rather than relying on haphazard searches of the internet. Fortunately, several organisations are now working to bring relevant evidence together into online resources. These include the following:

---

The Active Learning Network for Accountability and Performance (ALNAP) is hosted by the Overseas Development Institute (ODI) in London, UK, with a website containing more than 15,000 resources including ALNAP publications relevant to evidence and details of several thousand evaluations.

---

The foundations of Evidence Aid were laid down within Cochrane following the Indian Ocean tsunami of 2004, when the full text of several dozen Cochrane Reviews relevant to disaster response were made freely available online. It was established as a charity in 2015 and has now collated several hundred systematic reviews that are all free to view from its website ([www.evidenceaid.org](http://www.evidenceaid.org)), including, but not limited to, bundles of Special Collections of reviews relevant to windstorms, earthquakes, Ebola, the health of refugees and asylum seekers, post-traumatic stress disorder, and prevention of acute malnutrition in emergencies and humanitarian crises.

---

The International Initiative for Impact Evaluation (3ie) was established in 2008 and now offers four searchable databases online ([www.3ieimpact.org](http://www.3ieimpact.org)). Two of these, the 3ie Database of Systematic Reviews and the Database of Impact Evaluations catalogue evidence of the effectiveness of interventions in the humanitarian sector. These databases also include

systematic reviews and impact evaluations on the broader landscape of international development, many of which have relevance to interventions in emergency situations.

---

**We are maintaining an up-to-date fuller list of these types of resources online, at: [www.evidenceaid.org/online-collections-of-research-for-the-humanitarian-sector](http://www.evidenceaid.org/online-collections-of-research-for-the-humanitarian-sector)**

The list provides a wide range of online research resources, many of which are free and easy to access. These should be useful to any policy maker, NGO or frontline professional in the humanitarian sector, providing easy access to reliable, high quality evidence on the effectiveness of interventions. If you would like to suggest additional resources for this list, please contact Evidence Aid:

**[info@evidenceaid.org](mailto:info@evidenceaid.org)**

## CONCLUSION

**In conclusion to this guide on the use of evidence in the humanitarian sector, we encourage you to take advantage of the freely available, accessible and actionable summaries of research, such as the systematic reviews contained on the websites we have listed. This will help you to move quickly to sources of evidence that should help inform your policy and practice.**

## Key messages for Section D

✓ Research evidence of relevance to the humanitarian sector is scattered across tens of thousands of reports spread across thousands of journals, books and websites.

✓ Searching for this evidence may require help from an information specialist and designing the search needs to consider carefully what is being looked for, the sources to search, to terms to use in the search and any restrictions relating to language or time period.

✓ Systematic reviews provide summaries and synthesis of research evidence.

✓ Several online repositories of research are available, improving access to systematic reviews and other types of research evidence.

You might find it more efficient and easier to use trusted repositories of research and systematic reviews.

# Endnotes

- <sup>1</sup> Collins S, Sadler K. (2002) Outpatient care for severely malnourished children in emergency relief programmes: a retrospective cohort study. *Lancet* 360:1824-30.
- <sup>2</sup> WHO, World Food Programme, UN System Standing Committee on Nutrition, and UNICEF. (2007) Community-based management of severe acute malnutrition: a joint statement by the WHO, World Food Programme, UN System Standing Committee on Nutrition and UNICEF. New York: UNICEF.
- <sup>3</sup> Price AI, Djulbegovic B. (2017) What does evidence mean? Most languages translate “evidence” into “proof”. *Journal of Evaluation in Clinical Practice* 23(5):971-3.
- <sup>4</sup> Sackett D, et al. (1996) Evidence based medicine: what it is and what it isn't. *British Medical Journal* 312:71. See [www.bmj.com/content/312/7023/71](http://www.bmj.com/content/312/7023/71).
- <sup>5</sup> The definition of evidence-based medicine came at a time when most medical decision making was based on experience, authority and eminence. Medical practice was not informed by the best available scientific evidence. Some commentators and researchers have argued that social policy is in the same place as medicine was 20 or 30 years ago, namely that authority, rather than research evidence, dominates decision making.
- <sup>6</sup> Allen C, et al. (2016) Evidence Aid. *Oxford Public Health* August: 51-54; Blanchet K, et al. (2017) Evidence on public health interventions in humanitarian crises. *Lancet* 390:2287-96; and Christoplos I, et al. (2017) Strengthening the quality of evidence in humanitarian evaluations. ALNAP Method Note. London: ALNAP/ODI (see [www.alnap.org/system/files/content/resource/files/main/alnap-eha-method-note-5-2017.pdf](http://www.alnap.org/system/files/content/resource/files/main/alnap-eha-method-note-5-2017.pdf)).
- <sup>7</sup> Knox Clarke P, Darcy J. (2014) Insufficient evidence? The quality and use of evidence in humanitarian action. ALNAP Study. London: ALNAP/ODI. See [www.alnap.org/system/files/content/resource/files/main/alnap-study-evidence.pdf](http://www.alnap.org/system/files/content/resource/files/main/alnap-study-evidence.pdf).
- <sup>8</sup> [www.oxforddictionaries.com/definition/english/evidence](http://www.oxforddictionaries.com/definition/english/evidence).
- <sup>9</sup> Nutley S, et al. (2013) What Counts as Good Evidence? London, UK: Alliance for Useful Evidence.
- <sup>10</sup> Hallam A, Bonino F. (2013) Using Evaluation for a Change: Insights from humanitarian practitioners. ALNAP Study. London: ALNAP/ODI. See [www.alnap.org/system/files/content/resource/files/main/alnap-study-using-evaluation-for-a-change.pdf](http://www.alnap.org/system/files/content/resource/files/main/alnap-study-using-evaluation-for-a-change.pdf).
- <sup>11</sup> Ariely D. (2009) Predictably Irrational: The Hidden Forces that Shape Our Decisions. London, UK: HarperCollins.
- <sup>12</sup> HM Treasury. (2011) The Green Book: Appraisal and Evaluation in Central Government. London, UK: HM Treasury.
- <sup>13</sup> Christoplos I. (2006) Links between Relief, Rehabilitation and Development in the Tsunami Response: A Synthesis of Initial Findings. Stockholm, Sweden: Swedish International Development Cooperation Agency.
- <sup>14</sup> Zsombok CE, Klein G, (editors). (2014) Naturalistic decision making. New York, USA: Psychology Press.
- <sup>15</sup> Kahneman D, et al. (2009) Conditions for intuitive expertise: a failure to disagree. *American Psychology* 64:515-26. See [www.ncbi.nlm.nih.gov/pubmed/19739881](http://www.ncbi.nlm.nih.gov/pubmed/19739881).
- <sup>16</sup> These estimates come from Cash Learning Partnership (CaLP). (2018) The State of the World's Cash Report. Oxford: CaLP; Doing Cash Differently: Report of the High Level Panel on Cash Transfers. (2015); and Development Initiatives. (2017) Global Humanitarian Assistance Report.
- <sup>17</sup> ODI, Development Initiatives. (2016) Counting Cash: Tracking Humanitarian Expenditure on Cash-Based Programming.
- <sup>18</sup> Oxfam. (2006) Good Practice Review 11: Cash Transfer Programming in Emergencies. Oxford: Oxfam.

- <sup>19</sup> Jackson R (Save the Children UK), Kukrety N (Oxfam GB) (2012) Institutionalising cash transfer programming. See <https://odihpn.org/magazine/institutionalising-cash-transfer-programming/>.
- <sup>20</sup> See [www.humanitarianresponse.info/en/programme-cycle/space](http://www.humanitarianresponse.info/en/programme-cycle/space).
- <sup>21</sup> Discussion of the challenges of making decisions following the rare circumstances of a major radiation emergency is available in Carr Z, et al. (2016) Using the GRADE approach to support the development of recommendations for public health interventions in radiation emergencies. *Radiation Protection Dosimetry* 171:144-55; and Ohtsuru A, et al. (2015) Nuclear disasters and health: lessons learned, challenges, and proposals. *Lancet* 386: 489-97.
- <sup>22</sup> Bradley DT, et al. (2014) The effectiveness of disaster risk communication: a systematic review of intervention studies. *PLOS Currents Disasters* August 22; Edition 1.
- <sup>23</sup> A discussion of the importance of paying attention to how donors gather, use and share evidence and information is available in Obrecht A. (2017) Using Evidence to Allocate Humanitarian Resources: Challenges and Opportunities. ALNAP Working Paper. London: ALNAP/ODI. See <https://reliefweb.int/sites/reliefweb.int/files/resources/alnap-eaar-resource-allocation-2017.pdf>.
- <sup>24</sup> Nesta. (2013) Understand how innovation works. Video available at [www.nesta.org.uk/resources/understand-how-innovation-works](http://www.nesta.org.uk/resources/understand-how-innovation-works).
- <sup>25</sup> Obrecht A, Warner AT. (2016) More than just luck: Innovation in humanitarian action. HIF/ALNAP Study. London: ALNAP/ODI.
- <sup>26</sup> Jones G, et al. (2003) How many child deaths can we prevent this year? *Lancet* 362:65-71. See [https://linkinghub.elsevier.com/retrieve/pii/S0140-6736\(03\)13811-1](https://linkinghub.elsevier.com/retrieve/pii/S0140-6736(03)13811-1).
- <sup>27</sup> Nyhan B, Reifler J. (2010) When corrections fail: The persistence of political misperceptions. *Political Behavior* 32:303-30.
- <sup>28</sup> White H. (2009) Theory-based impact evaluation: principles and practice, working paper 3. Delhi, India: International Initiative for Impact Evaluation, page 4.
- <sup>29</sup> Waddington H, et al. (2009) Water, Sanitation and hygiene interventions to combat childhood diarrhoea in developing countries: a systematic review. Delhi, India: International Initiative for Impact Evaluation.
- <sup>30</sup> Nesta and TSIP. (2014) Guidance for Developing a Theory of Change for Your Programme. See [www.nesta.org.uk/sites/default/files/theory\\_of\\_change\\_guidance\\_for\\_applicants\\_.pdf](http://www.nesta.org.uk/sites/default/files/theory_of_change_guidance_for_applicants_.pdf).
- <sup>31</sup> Stern E. (2015) Impact Evaluation: A Design Guide for Commissioners and Managers of International Development Evaluations in the Voluntary and Community Sector. London, UK: Big Lottery Fund, Bond, Comic Relief and the Department for International Development.
- <sup>32</sup> Gerdin M, et al. (2014) Optimal evidence in difficult settings: improving health interventions and decision making in disasters. *PLoS Medicine* 11(4): e1001632.
- <sup>33</sup> Mulgan G. (2015) The six Ws: a formula for what works. London: Nesta. See [www.nesta.org.uk/blog/six-ws-formula-what-works](http://www.nesta.org.uk/blog/six-ws-formula-what-works).
- <sup>34</sup> [www.ebola-anthropology.net/about-the-network](http://www.ebola-anthropology.net/about-the-network).
- <sup>35</sup> Faye SL. (2015) L'“exceptionnalité” d'Ebola et les “réticences” populaires en Guinée-Conakry. Réflexions à partir d'une approche d'anthropologie symétrique. *Anthropologie & Santé*. See <https://journals.openedition.org/anthropologiesante/1796>.
- <sup>36</sup> <http://pubman.mpdl.mpg.de/pubman/item/escidoc:2096578/component/escidoc:2103624/AAA-Ebola-Report-1.pdf>.
- <sup>37</sup> Fairhead J. (2016) Understanding Social Resistance to the Ebola Response in the Forest Region of the Republic of Guinea: An Anthropological Perspective. *African Studies Review* 59:7-31. doi:10.1017/asr.2016.87.
- <sup>38</sup> Abramowitz S, et al. (2015) Social science intelligence in the global Ebola response. *Lancet* 385:330.
- <sup>39</sup> Petticrew M, et al. (2003) Evidence, hierarchies and typologies: horses for courses. *Journal of Epidemiology and Community Health* 57:527-9.

- <sup>40</sup> An authoritative and exhaustive list of social science research frameworks and methods is available in Luff R, et al. (2015) Review of the Typology of Research Methods within the Social Sciences. London, UK: ESRC/ National Centre for Research Methods. See <http://eprints.ncrm.ac.uk/3721>.
- <sup>41</sup> Department for International Development. (2014) How to Note: Assessing the Strength of Evidence. See [www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/291982/HTN-strength-evidence-march2014.pdf](http://www.gov.uk/government/uploads/system/uploads/attachment_data/file/291982/HTN-strength-evidence-march2014.pdf).
- <sup>42</sup> Adapted from HM Treasury, DECC and DEFRA. (2012) Quality in policy impact evaluation: understanding the effects of policy from other influences. London, UK: HM Treasury/DEFRA/ DECC; Frost S, et al. (2006) The Evidence Guide; Using Research and Evaluation in Social Care and Allied Professions. London, UK: Barnardo's; Petticrew M, Roberts H. (2003) Evidence, hierarchies and typologies: horses for courses. *Journal of Epidemiology and Community Health*. 57: 527-9; and Stern E. (2015) Impact Evaluation; A Design Guide for Commissioners and Managers of International Development Evaluations in the Voluntary and Community Sector. London, UK: Big Lottery Fund, Bond, Comic Relief and the Department for International Development, Table 2, page 18.
- <sup>43</sup> Odgaard-Jensen J, et al. (2011) Randomisation to protect against selection bias in health care trials. *Cochrane Database of Systematic Reviews* (4):MR000012.
- <sup>44</sup> White H. (2013) An introduction to the use of randomised control trials to evaluate development interventions. *Journal of Development Effectiveness* 5(1):30-49.
- <sup>45</sup> See for example: Schulz KF, et al. (2010) CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *PLoS Medicine* 7(3):e1000251; Higgins JP, et al. (2011) The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 343:d5928.
- <sup>46</sup> Puri J, et al. (2015) What methods may be used in impact evaluations of humanitarian assistance? working paper 22, Delhi, India: International Initiative for Impact Evaluation. See [www.3ieimpact.org/media/filer\\_public/2014/12/08/wp\\_22\\_humanitarian\\_methods\\_working\\_paper-top.pdf](http://www.3ieimpact.org/media/filer_public/2014/12/08/wp_22_humanitarian_methods_working_paper-top.pdf).
- <sup>47</sup> Doocy S, Burnham G. (2006). Point-of-use water treatment and diarrhoea reduction in the emergency context: an effectiveness trial in Liberia. *Tropical Medicine and International Health* 11:1542-52.
- <sup>48</sup> White H, et al. (2014) Randomised Controlled Trials (RCTs). Methodological Briefs. Impact Evaluation No. 7. Florence, Italy: Unicef. See [www.unicef-irc.org/publications/pdf/brief\\_7\\_randomised\\_controlled\\_trials\\_eng.pdf](http://www.unicef-irc.org/publications/pdf/brief_7_randomised_controlled_trials_eng.pdf).
- <sup>49</sup> Gertler P. (2000) Final Report: The Impact of PROGRESA on Health. Washington, DC, USA: International Food Policy Research Institute (IFPRI). See [www.ifpri.org/publication/impact-progres-health](http://www.ifpri.org/publication/impact-progres-health).
- <sup>50</sup> Hainmueller J, et al. (2017) Catalyst or crown: does naturalization promote the long-term social integration of immigrants? *American Political Science Review* 111(2):256-76. doi:10.1017/S0003055416000745.
- <sup>51</sup> Bozorgmehr K, et al. (2015) Effect of restricting access to health care on health expenditures among asylum-seekers and refugees: A quasi-experimental study in Germany, 1994–2013. *PLoS ONE* 10: e0131483.
- <sup>52</sup> Rossi R, et al. (2016) Vaccination coverage cluster surveys in Middle Dreib - Akkar, Lebanon: comparison of vaccination coverage in children aged 12-59 months pre- and post-vaccination campaign. *PLoS ONE* 11(12): e0168145.
- <sup>53</sup> Alexander J, Bonino F (2015). A discussion on designs, approaches and examples, ALNAP Discussion Series Improving the quality of EHA evidence, Method Note 4, January 2015. See [www.alnap.org/system/files/content/resource/files/main/alnap-eha-method-note-addressing-causation-jan2015.pdf](http://www.alnap.org/system/files/content/resource/files/main/alnap-eha-method-note-addressing-causation-jan2015.pdf).
- <sup>54</sup> Waddington H, et al. (2017) Quasi-experimental study designs series paper 6: risk of bias assessment. *Journal of Clinical Epidemiology* 89:43-52; and Bärnighausen T, et al. (2017) Quasi-experimental study designs series—paper 7: assessing the assumptions. *Journal of Clinical Epidemiology* 89:53-66.

- <sup>55</sup> Nielsen NS, et al. (2013). The Contribution of Food Assistance to Durable Solutions in Protracted Refugee Situations; its impact and role in Bangladesh: A Mixed Method Impact Evaluation, Volume I-Evaluation Report, Geneva, Switzerland: World Food Program/UNHCR.
- <sup>56</sup> Carter R. (2012) Helpdesk Research Report: Theory-based evaluation approach. University of Birmingham, UK: Birmingham: Governance and Social Development Resource Centre. See [www.gsdr.org/docs/open/hdq872.pdf](http://www.gsdr.org/docs/open/hdq872.pdf).
- <sup>57</sup> Westhorp G. (2014) Realist Evaluation: An Introduction. London, UK: Methods Lab, Overseas Development Institute. See [www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/9138.pdf](http://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/9138.pdf).
- <sup>58</sup> Mayne J. (2012) Contribution Analysis: Coming of Age? *Evaluation* 18(3):270-80.
- <sup>59</sup> For network analysis and process tracing see: Befani B, et al. (2014) Process tracing and contribution analysis: a combined approach to generative causal inference for impact evaluation. *IDS Bulletin* 45(6):17-36.
- <sup>60</sup> Baptist C, et al. (2015) Coffey How To Guide: Qualitative Comparative Analysis - A Rigorous Qualitative Method for Assessing Impact. See [www.coffey.com/assets/Ingenuity/Qualitative-Comparative-Analysis-June-2015.pdf](http://www.coffey.com/assets/Ingenuity/Qualitative-Comparative-Analysis-June-2015.pdf).
- <sup>61</sup> Stern E, et al. (2012) Broadening the range of designs and methods for impact evaluations, London, UK: Department for International Development.
- <sup>62</sup> For example, see Carol Weiss's guide: Weiss C. (1998) *Evaluation: methods for studying programs and policies*. New York, USA: Prentice Hall.
- <sup>63</sup> Ioannidis JA. (2005) Contraindications and Initially Stronger Effects in Highly Cited Clinical Research. *Journal of the American Medical Association*. 294(2):218-28.
- <sup>64</sup> For a systematic review of publication bias see Hopewell S, et al. (2009) Publication bias in clinical trials due to statistical significance or direction of trial results. *Cochrane Database of Systematic Reviews* (1):MR000006. doi: 10.1002/14651858.MR000006.pub3.
- <sup>65</sup> Franco A, et al. (2014) Social science. Publication bias in the social sciences: unlocking the file drawer. *Science* 345:1502-5.
- <sup>66</sup> Sense about Science. (2006) I don't know what to believe: Making sense of science stories. See [www.senseaboutscience.org/resources.php/16/i-dont-know-what-to-believe](http://www.senseaboutscience.org/resources.php/16/i-dont-know-what-to-believe).
- <sup>67</sup> See for instance, the supplement to the Magenta Guide: HM Treasury, DECC and DEFRA (2012) Quality in policy impact evaluation; understanding the effects of policy from other influences. The guidance shows how 'higher quality research designs can help meet the challenge of attributing measured outcomes to the policy in question (as opposed to other influences), whereas lower quality designs reduce confidence in whether it was the policy that achieved those outcomes' (page 5).
- <sup>68</sup> <http://gradeworkinggroup.org>.
- <sup>69</sup> Farrington RP, et al. The Maryland Scientific Methods Scale, In: Farrington DP, et al. *Evidence-based crime prevention*. London, UK: Routledge, 2002 chapter 2.
- <sup>70</sup> For a model of different approaches on quality that includes the four dimensions of (1) methodological quality (2) quality in reporting, (3) appropriateness and (4) relevance to policy and practice; see Boaz A, Ashby D. (2003) Fit for purpose? Assessing research quality for evidence based policy and practice. Working Paper 11. London, UK: ESRC UK Centre for Evidence Based Policy and Practice.
- <sup>71</sup> Callaway E. (2011) Report finds massive fraud at Dutch universities. *Nature* 479(15). See [www.nature.com/news/2011/111101/full/479015a.html](http://www.nature.com/news/2011/111101/full/479015a.html).
- <sup>72</sup> Boaz A, et al. (2003) Fit for purpose? Assessing research quality for evidence based policy and practice. Working Paper 11. London, UK: ESRC UK Centre for Evidence Based Policy and Practice.
- <sup>73</sup> Qualitative research explores and tries to understand people's beliefs, experiences, attitudes, behaviour and interactions. It generates non-numerical data which might be gathered through, for example in-depth interviews, focus groups, documentary analysis and participant observation.

- <sup>74</sup> Spencer L, et al. (2002) *Quality in Qualitative Evaluation: A framework for assessing research evidence*. London, UK: Cabinet Office; CASP. (2018) *CASP Checklist for Qualitative Research*. Oxford, UK: Critical Appraisal Skills Programme; National Institute for Health and Clinical Excellence. (2012) *The guidelines manual: Appendix H: Methodology checklist: qualitative studies*. London, UK: National Institute for Health and Clinical Excellence; and O'Brien BC, et al. (2014) Standards for reporting qualitative research: a synthesis of recommendations. *Academic Medicine* 89: 1245-51.
- 
- <sup>75</sup> Whitty CJM. (2015) What makes an academic paper useful for health policy? *BMC Medicine* 13:301.
- 
- <sup>76</sup> See, for example, the discussion of the importance of using existing evidence when designing new studies in Clarke M. (2004) Doing new research? Don't forget the old: nobody should do a trial without reviewing what is known. *PLoS Medicine* 1: 100-2; and the history of systematic reviews in Chalmers I, et al. (2002) A brief history of research synthesis. *Evaluation and the Health Professions* 25: 12-37; and Clarke M. (2016) History of evidence synthesis to assess treatment effects: personal reflections on something that is very much alive. *JLL Bulletin: Commentaries on the history of treatment evaluation. Journal of the Royal Society of Medicine* 109: 154-63.
- 
- <sup>77</sup> Allen C. (2014) A resource for those preparing for and responding to natural disasters, humanitarian crises, and major health care emergencies. *Journal of Evidence-Based Medicine* 7: 234-7; and Gurevitch J, et al. (2018) Meta-analysis and the science of research synthesis. *Nature* 555: 175-82.
- 
- <sup>78</sup> Oxman AD. (1994) Checklists for review articles. *BMJ* 309: 648-51.
- 
- <sup>79</sup> For a critique of applying research syntheses into policy, see Pawson R. (2001) *Evidence Based Policy: In Search of a Method*. Working Paper 3. London, UK: ESRC UK Centre for Evidence Based Policy and Practice.
- 
- <sup>80</sup> Donnelly CA, et al. (2018) Four principles for synthesizing evidence. *Nature* 558: 361-4.
- 
- <sup>81</sup> Lewis S, Clarke M. (2001) Forest plots: trying to see the wood and the trees. *BMJ* 322:1479-80.
- 
- <sup>82</sup> Bangpan M, et al. (2017) The impact of mental health and psychosocial support programmes for populations affected by humanitarian emergencies. Oxford, UK: Oxfam.
- 
- <sup>83</sup> Ganann R, et al. (2010) Expediting systematic reviews: methods and implications of rapid reviews. *Implementation science* 5:56; and Tricco AC, et al. (2015) A scoping review of rapid review methods. *BMC Medicine* 13: 224.
- 
- <sup>84</sup> Begley CG, et al. (2012) Drug development: Raise standards for preclinical cancer research. *Nature* 483: 531-3.
- 
- <sup>85</sup> A review of research that had included one or more cumulative meta-analysis found many examples showing that stable results (beneficial, harmful and neutral) would have been seen had a meta-analysis of existing research been done before a new randomised trial began, which would have led to earlier uptake of effective interventions: Clarke M, et al. (2014) Accumulating research: a systematic account of how cumulative meta-analyses would have provided knowledge, improved health, reduced harm and saved resources. *PLoS ONE* 9(7): e102670.
- 
- <sup>86</sup> Chalmers I, et al. (2014) How to increase value and reduce waste when research priorities are set. *Lancet* 383: 156-65.
- 
- <sup>87</sup> Shea BJ, et al. (2007) Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Medical Research Methodology* 7:10; and Shea BJ, et al. (2017) AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of health care interventions, or both. *BMJ* 358:j4008.
- 
- <sup>88</sup> Smith V, et al. (2011) Methodology in conducting a systematic review of systematic reviews of health care interventions. *BMC Medical Research Methodology* 11:15.
- 
- <sup>89</sup> See, for example: Brennan RJ, et al. (2005) Rapid health assessment in Aceh Jaya district, Indonesia, following the December 26 tsunami. *Emergency Medicine Australasia* 17: 341e350; and Beebe J. (2014) *Rapid Qualitative Inquiry: a Field Guide to Team-based Assessment*, second edition. Lanham, Maryland, USA: Rowman & Littlefield.
-

Karl Blanchet, Director of the Health in Humanitarian Crises  
Centre, London School of Hygiene & Tropical Medicine  
[Karl.Blanchet@lshtm.ac.uk](mailto:Karl.Blanchet@lshtm.ac.uk) | [crises.lshtm.ac.uk](http://crises.lshtm.ac.uk)

Mike Clarke, Research Director of Evidence Aid,  
Centre for Public Health, Queen's University Belfast  
[mclarke@qub.ac.uk](mailto:mclarke@qub.ac.uk) | [www.qub.ac.uk](http://www.qub.ac.uk)



Cover Photo © Patrick Brown/UNICEF/Panos Pictures