Impact evaluations of humanitarian assistance interventions:
A brief on scope and methods.
An executive summary

Understanding the impact of humanitarian assistance is an area where much work is needed... Linking impact measurement and accountability better to the funds agencies receive is a key recommendation of this review (Humanitarian Emergency Response Review, UK Government, March 2011).

The evidence base proving which humanitarian responses are most effective is extremely lacking. Investments must be made in the consolidation of evidence about what works in response to different kinds of needs in different contexts (The Use of Evidence in Humanitarian Decision Making, ACAPS Operational Learning Paper, January 2013).

I. Objectives of the study

Humanitarian crises are complex situations where the demand for aid has traditionally far exceeded its supply. The humanitarian assistance community has long asked for better evidence on how each dollar should be effectively spent. Impact evaluations of humanitarian assistance can help answer these questions and also respond to the increasing call to estimate the impact of humanitarian assistance and supplement the rich tradition for undertaking real-time and process evaluations in the sector.

Funded by DFID & USAID, led by 3ie and undertaken with the assistance of Evidence Aid, SIPRI (Stockholm International Peace Research Institute) and humanitarian assistance experts, the study aims to answer three main questions:

i) Have impact evaluations of humanitarian assistance been undertaken? If yes, in what sectors and sub-sectors have these been undertaken?¹

¹ For the purposes of the study, impact evaluations are defined as studies of measurable changes in outcomes and impact (and in some cases even outputs) that can be attributed in a causal way to the activities or interventions or a program undertaken in the context of humanitarian assistance.
ii) What methodological techniques can we use to address some of the important questions in this area while simultaneously considering the special circumstances and constraints of humanitarian assistance?

iii) In protracted emergencies and sudden-onset disasters, what studies can best help inform the programming and strategic planning needs of the humanitarian community?

II. Key findings for the scope of future study

A survey of almost 400 respondents, semi-structured interviews of 50 experts in the field, a review of strategy documents of main organizations in the sector and a review of repositories of evaluations showed:

- **Insufficient high-quality evidence**: High-quality evidence that can causally relate changes in the conditions of people and their outcomes to specific programmes and interventions undertaken in humanitarian assistance are clearly scarce. In an investigation of studies conducted since 2005, we found 39 studies that could be described as impact evaluations that used (implicit or explicit) comparison groups to measure attributable change. However these too were deficient in many ways: 29 had a theory of change but 23 did not show whether the choice of comparison groups was valid (i.e. did not have balance tests); 29 did not discuss the confidence with which their results were measured (i.e. did not undertake power analyses or show sample size calculations) and only five discussed ethical issues.

- **Sectors**: Using gap-maps of evidence, the study finds that most high-quality studies of humanitarian assistance, are in the area of health (and particularly mental health), nutrition and peace-building.

- **Timing**: Most existing impact evaluation studies examine changes in conditions and resilience once the affected area is in the recovery phase (there are approximately 27 studies that examine the results of peace building and conflict prevention). There are few studies of unanticipated disasters (four) – all of which examine recovery and resilience and few studies (six) of efforts of immediate relief.

- **A needs map**: A needs map drawn from interviews, strategy documents and interviews helped us to visualise main areas in which practitioners require additional evidence and research. In particular, more than 20% cited accountability, food security, protection, water and sanitation, and health and said that it was important to assess their impact not just on food security but also on nutrition, income and in the longer term on recovery and resilience. Education, humanitarian assistance as a whole, nutrition and logistics were
said to be important for study by 10-20% and less than 10% suggested emergency telecommunications and camp management.

III. Key findings for methods

There are many constraints that impact evaluations need to overcome in humanitarian situations, in addition to those that face studies in less complex and challenging situations. The robustness of studies can be especially compromised in the absence of baseline data and inability to plan for and construct counterfactuals. The need for speed of action and low predictability also usually means that little advance preparation is possible. Furthermore, most humanitarian situations have a multiplicity of actors and it is usually difficult to de-couple actions and outcomes. High-covariability or the fact that conflict and disasters don’t usually have clean boundaries mean that it is also difficult to find or establish comparable groups that can serve as counterfactuals in a scientifically robust and ethically sound way. Last but not least, there is a lack of impact evaluation experts in the humanitarian sector and a lack of humanitarian experts in the impact evaluation sector.

However, traditional evaluations that either monitor processes or assess if targets have been achieved are clearly insufficient for robust evidence by themselves. They are unable to examine unintended consequences; or to deal with a variety of biases, such as selection bias (i.e. areas targeted by humanitarian assistance area likely to have attributes that make them more or less likely to recover, compared to the average), non-random attrition (are unlikely to count people who either migrated as a result of the intervention or those who perished as a result of the disaster), and contamination bias (areas targeted by one actor are also likely to have other sources of assistance that may make it difficult to separate the different sources of changes); and are unable to measure the change that has occurred as a consequence of their action.

Clearly, new ways must be forged to combine the strengths of methods and traditions that exist in the sector so that these can be used to enhance evidence, while responding to increased demand for accountability in the context of rapid-onset and protracted crises. Additional questions that impact evaluations can help answer are:

- How much was the change in conditions a result of the program? Was the affected population able to recover to their pre-disaster levels? Are we 'building back better'?
- How much of the program/intervention should be delivered, with what frequency?
- What is the best way to deliver an assistance package? What difference did it make? Can it be delivered in a more cost-effective manner?
- How much difference did my agency make?
- Were some groups better off as a result of the program compared to others?
Methods for undertaking impact evaluations should therefore address concerns for costs, speed, multiplicity of actors, absence of data and the ethics of responding speedily to disrupted communities while maintaining the need to address the needs of the most vulnerable.

In six case studies in the paper, we discuss possible methods to address these concerns in different humanitarian contexts that range from unanticipated natural disaster related emergencies to protracted crises. We discuss a variety of situations and present examples of possible methods that may be used to answer important questions. These situations include understanding the effectiveness of ready-to-use supplementary food in post-conflict Sri Lanka, the effectiveness of trust building interventions in post-conflict areas in Kyrgyzstan, education for women in earthquake affected areas, the effectiveness of cash-based flood relief on nutrition and food consumption in a complex emergency in Pakistan, the relative effectiveness of food coupons versus cash for internally displaced populations in a protracted emergency in northern DRC, post-typhoon emergency assistance for pregnant women in the Philippines and an intervention to improve adherence to tuberculosis therapy in South Sudan.

The main findings from these case studies are:

**Ethics**: In our review and discussions of case studies we use the ‘no-harm principle’. Methods used for assessing and measuring changes in outcomes all use approaches that turn the challenge of limited resources and the inability to cover all affected populations at once, into an opportunity. The use of phased roll-outs of interventions and rolling out programs with small changes while keeping the basic package the same (also called a factorial design) are discussed in relation to the effectiveness of nutrition related interventions and cash transfers. In other cases where it is not possible to randomly assign packages for an impact evaluation ex-ante, we discuss the use of ex-post quasi-experimental designs (such as propensity score matching and regression discontinuity design) that can be used to assess the effect of SMSes on TB adherence and programs to increase the uptake of iron tablets.

**Data availability**: In all cases, we discuss the use of other available data that may be used creatively in the context of both unanticipated disasters and for protracted crises. So, in the case of examining food distribution in Sri Lanka, training in Kyrgyzstan and post-earthquake education in Pakistan, we discuss the use of the Living Standards Measurement Surveys (LSMS). Additionally, in the case of post-

---

2 Humanitarian Accountability Project has been on a long quest to demonstrate the ‘business case’ for beneficiary accountability but was never able to provide evidence that being accountable to beneficiaries using HAP’s standards led to improved impact.

3 The approach to be tested may significantly improve but will not worsen outcomes for emergency relief recipients.
typhoon Philippines, the DHS (Demographic and Health Surveys) may be used to understand minimum detectable effect sizes, calculate appropriate sample sizes and create a baseline. In other cases, such as post-Typhoon Philippines we discuss the use of satellite images and Geographic Information Systems and the use of mobile phones for easy and cheap data collection in the case of TB adherence in South Sudan.

**Costs and speed:** The cost of robust impact evaluations that help to clearly understand and measure changes in outcomes is more than that of no evaluations at all. However, agencies regularly undertake many evaluations whether these be process evaluations, monitoring or real-time evaluations. Good theory-based impact evaluations contain all of these but go a step further in assessing attributable changes. Rapid impact evaluations that use individual assignment and are quick and less costly have been conducted for example, to study malaria interventions, to assess ready to use supplementary food and to determine the effectiveness of approaches to increase voluntary medical male circumcision in mostly development contexts. Customizing these types of study to the needs to the humanitarian community is a possibility and should be discussed widely. In the case studies, most examples use individual assignment and lend themselves to rapid evaluations.

**Sample sizes:** Another concern for most practitioners is the size and cost of data collection. Using the six case studies, we show that it is usually not necessary to collect data on all beneficiaries. Surveys that inform impact evaluations usually need to be conducted only on a subset of the affected population. Using case studies we show that depending on the outcomes, the range of sample sizes required for undertaking good impact studies varies from 300 individuals (for food distribution and its effect on haemoglobin in Sri Lanka), to 690 households in 30 clusters (for comparing cash versus food coupons in DRC), to 2000 households in 30 clusters (to understand factors affecting trust in Kyrgyzstan). Thus, the cost of data collection may be neither prohibitive or so widespread as to interfere with field operations.

**Unintended consequences and vulnerable groups:** An important concern for many respondents was understanding the impact on vulnerable groups such as women, children, people living in remote areas and chronically poor populations. Investigating the differential distribution and uptake of humanitarian assistance amongst these groups is clearly important. It is possible to design impact evaluations that can measure this differential impact but also assess other unintended consequence with the help of good theories of change. Another important question raised by respondents is the need for a better understanding and estimating the effect of protocols for increased coordination amongst humanitarian agencies. We present some hypothetical solutions for understanding these in the case of distribution of nutrition packages in Sri Lanka for example, where a multitude of agencies work. But we advocate for more thought and discussion especially with respect to what outcomes may be most useful to investigate.
IV. Conclusions

Since 2005 more than $90 billion has been spent on humanitarian assistance. Fortunately several agencies around the world are taking the lead and showing that institutional constraints, capacity bottlenecks and concerns about image-risk that are traditionally associated with impact evaluations can be transcended. Yet clearly very few impact evaluations are being conducted. This study represents a first assessment to our knowledge of the scope and possible methods that may be used to undertake high quality impact evaluations of humanitarian assistance. Three critical steps can help take this effort forward. Firstly, agencies can pilot impact evaluations in areas that are relevant to them and help demonstrate feasibility and practice. Secondly, a dialogue on the priorities and feasibility of these types of studies of humanitarian assistance can help remove some of the myths that surround the use and planning of impact evaluations. Finally, clearly earmarked resources can contribute to building a critical body of evidence that can inform programming and strategy in humanitarian assistance clearly and consistently.

To know more, to volunteer to expert review the draft version of this study or to understand how your organization may participate in this effort, kindly contact us.

Dr Howard White: Executive Director, 3ie. ed@3ieimpact.org
Dr Jyotsna (Jo) Puri: Head of evaluation and Deputy Executive Director, 3ie. Jpuri@3ieimpact.org