

Evaluating Evaluations and Assessing Assessments: NGO Views on Humanitarian Interventions

Steve Hansch, Ph.D.

Center for Disaster and Humanitarian Assistance Medicine
Uniformed Services University Health Sciences
and the Henry Jackson Foundation.
Maryland, U.S.A.
e-mail: shansch@bellatlantic.net

Steve Hansch is special projects officer at the Center for Disaster and Humanitarian Assistance Medicine, based in the Uniformed Services University of the Health Sciences in Maryland. Having studied epidemiology and nutrition at Stanford, Boston University and Tufts University, he has spent his career working mostly with NGOs, including the Refugee Policy Group, IRC, CARE, the Congressional Hunger Center, and, currently, the Cuny Center. He has conducted evaluations or implemented programs in some 130 refugee camps or disaster areas in roughly 40 countries.

The goal of this paper is to review how humanitarian NGOs assess and evaluate needs in emergencies and their own performance. After a summary of methods and techniques, this paper will summarize a few key observations about the impact of emergencies. Much of the later part of the paper will focus on the limitations of our techniques and confusions and myths.

The scope of this review includes the broad range of humanitarian endeavor, which Canada's recently-retired foreign minister, Lloyd Axworthy, demonstrated through his interest in what he called "Human security", which encompasses livelihood, famine, natural disasters, landmines, arms trade, sanctions, failed states, conflict and complex emergencies.

TERMINOLOGY

In the nongovernmental world we tend to use the word "Assessment" primarily to refer to the act of measuring needs within a population who have been affected by an emergency. In contrast, we tend to use the word evaluation for something that happens afterwards, to look back at what we "did," with the result of drawing lessons learned and insights into whether projects should be continued, expanded in size or replicated elsewhere.

NGOs think and plan in terms of the "project." NGOs assess in order to help design a project, and evaluations typically evaluate one project. Each project has its own unique time-frame, personality, mix of activities, obligations, donors, goals and objectives. NGOs sell the project, fundraise for it and hire staff specific to implementing the project.

In effect, NGOs triage between project options in emergencies. Individuals are rarely triaged, but whole communities are. When an NGO reviews its scarce resources and decides to just focus on a water project, instead, for example, of doing a lot of food-aid, search-

rescue, and medical care, there is an intrinsic triage occurring between interventions. NGOs rarely can or choose to take on all sectors at once.

In thinking or talking about a project, NGOs use the analytic tool called the Logframe, which indeed is a requirement imposed by US, Canadian and other government donors. The logframe simply helps us to be more clear about what are inputs, outputs, actions taken, objectives and goals.

NGOs talk a great deal about “targeting” (as do aid donors), which follows from the recognition that they can not reach everyone perfectly with all desirable services. Choices need to be made, resources focused. Often, targeting is the NGOs way of recognizing that certain groups are more vulnerable than others, or are less likely to be capable of availing themselves of general aid programs, so special aid interventions are stood up specific for these sub-populations.

The Figure 1 depicts the life-cycle of a particular project, following a stress event (which could be a refugee flow or earthquake or violent conflict). Assessment measures the range and severity of needs, comparing regions, comparing populations. Appraisal is intimately connected with assessment. Appraisal refers to how the NGO determines what can be done. To understand what are feasible interventions, the NGO examines local “counterparts,” local groups and coping abilities. NGOs also measure local physical capacities – warehouses, food stocks, electrical power, potential (even if untapped) water sources. And NGOs always appraise the prospective problems they will encounter when moving in – both the physical capability of roads, airports, ports and barriers they might encounter from heavy rainfall (washing out roads) or barriers from government or social red-tape.

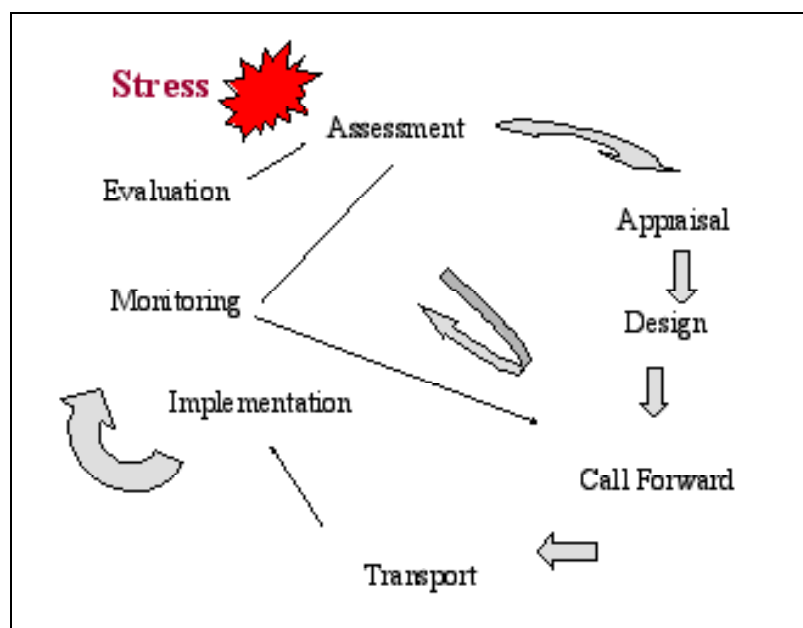


Figure 1: The project life-cycle is triggered by a stress event of some type.

As an example, in the case of food aid, NGOs intending to use USG food aid (under Public Law 480), are required to appraise (or project) the likelihood that new food aid will disrupt or harm local food aid markets, or displace international (and US) commercial food sales to that country.

After weeks or months of assessment and appraisal, the NGO then “designs” a project, which is another word for writing a detailed plan. Planning establishes a document which proposes the project, which is then circulated to potential donors. Then resources are raised, food aid is “called forward,” people and goods are transported to theater, and the project begins implementation.

Then when the project is active, NGOs “monitor” its progress. Monitoring refers both to performance and impact as well as tracking the movement of goods along the logistics pipeline.

KEY METRICS

Despite the mounds of verbiage and papers and data, NGOs and USAID invariably fall back on the use of a few indices, particularly in assessment.

Deaths are universally defined. Hindus in India, native Americans in Peru and Kosovars all have a standard definition of when a person is dead, so the “Crude Mortality Rate” tends to be a readily available and standardized metric that proves fundamental to assessing severity of need. Based on insights by the Centers for Disease Control and Prevention (CDC), later picked up by OFDA and UNHCR, NGOs often use the idea of “excess mortality” which is defined as the crude mortality rate subtracting the “baseline” mortality rate. In other words, we identify the difference between the observed (new) mortality and the mortality rate seen during the same season in a prior non-emergency year. In fact, disaster professionals define a disaster as a situation where the crude mortality rate rises above .5 deaths per 10,000 people per day.

The second most-used index is the “Rate of Malnutrition,” that is the proportion of children in the overall population who are malnourished. Most commonly, malnutrition is measured by comparing the weight of a child to the child’s height. We have international standards for definitions of malnutrition at different ages, weights and heights. Malnutrition is defined universally, without regard to race, culture, geography.

The value of rates of malnutrition is that the source of the information are people themselves, it represents empirical downstream impact. When an NGO identifies a rate of malnutrition of 30%, it signals an emergency with high likelihood of excess mortality.

Traditionally we have used food balance sheets as an indirect gauge of food problems. It is a macro calculation, depending on broad measures of country food stocks, food production and trade. It does not tell us about what is happening among people. We can measure the extent of aid needed by using the variable “Food Deficit” defined as food needs (MT) - food available (MT).

Finally, the fourth main variable we use is the number of persons displaced. It’s used in part because it’s a number that is readily generated in the field. People count refugees, or count people fleeing a village. Whether its people fleeing a natural disaster, a war or a famine, we know that flight itself is a risk factor, since people incur heightened rates of malnutrition in any high-density camp environment.

METHODS OF MEASUREMENT

There is a rich literature of tools and techniques for field collection of data (less, on data analysis or synthesis). Many of the techniques involve rapid gathering of qualitative data, through interviews, focus group discussions, village mapping, and direct observation. Among the most valuable types of data we collect are from existing indices such as price quotes for staple foods or hospital entries.

For malnutrition rates and many other critical variables, our emphasis is on finding information that represents a large population. In order to achieve representative-ness, we use random sampling methods. Because emergency affected, displaced (and refugee) populations are irregularly organized and imperfectly registered or counted, we have learned not to trust any rolls or lists. So our main tool is a two-tier random sampling, which we call two-stage cluster sampling. We randomize the regions of an area or sectors of a camp. So, we randomly select, say, 30 bite-size areas out of the overall universe (of say 100 regions). Then within each area we randomly select homesteads (or tents or huts), and once selected we have to measure every child in that dwelling, even if it means that we have to return to the hut at repeated times to snag them.

It would be useful, if there were more time, to explain that we use different tools, and slightly different metrics for different purposes. We might, for example use different measures of malnutrition for each of these separate tasks:

- Rapid comparison of villages.
- Screening a population, for example on entry into a camp;
- Repeat checking of individuals to evaluate the accuracy of our definitions and measurement tools;
- Tracking an individual child on recovery from malnutrition;
- Monitoring the progress of a population (e.g., how a refugee camp's nutritional situation, its mean and spread, changes over time);
- Monitoring the efficacy of a particular program (for example a targeted supplementary feeding project);
- Determining post-hoc the cost-efficiency (and therefore whether to repeat) our interventions.

After looking in frustration at mounds of poignant and significant data reported by agencies from many points in Somalia, the Centers for Disease Control and Prevention (CDC) team based in Mogadishu, made a plea to NGOs everywhere to adhere to a simple standard for reporting measures in emergencies. They asked that NGOs always add the following information about their data to any reporting:

1. Specify the population (as a bounded set) in question, whom the observed group represents e.g., is it 5,000 or 5 million, urban or rural.

2. Specify the timeframes in question (two month retrospective or two days?)
3. Specify the method of observation (recruitment, or sampling or testing, or self-selection) thoughts.

EMPIRICAL VIEWS OF THE IMPACT OF DIFFERENT EMERGENCIES

In humanitarian crises, NGOs are concerned to respond to the main problems: deaths, disability, suffering — in the right ways at the right times, and to learn what interventions prove effective at these different points in time

Reviews during the past fifteen years have revealed that the death rates attributable to emergencies are well predicted by the rate of malnutrition in the population. This is an ecological study: the rate of population malnutrition rate per refugee camp is well correlated with the simultaneous mortality rate in the same camps.

It also turns out that the bell curve (or binomial-like distribution) of population nutrition (measured by weight-for-height; but equally true for any anthropometric measure we use) remains the same shape regardless of the emergency. It doesn't spread or become binomial. The whole curve moves south (left, worse) under stress.

Excess mortality patterns differ dramatically and predictably by the type of disaster. As a generality, the shorter the onset of the crisis, the less the total mortality. Most people can run away from most natural disasters. They do not lead to epidemics or famine.

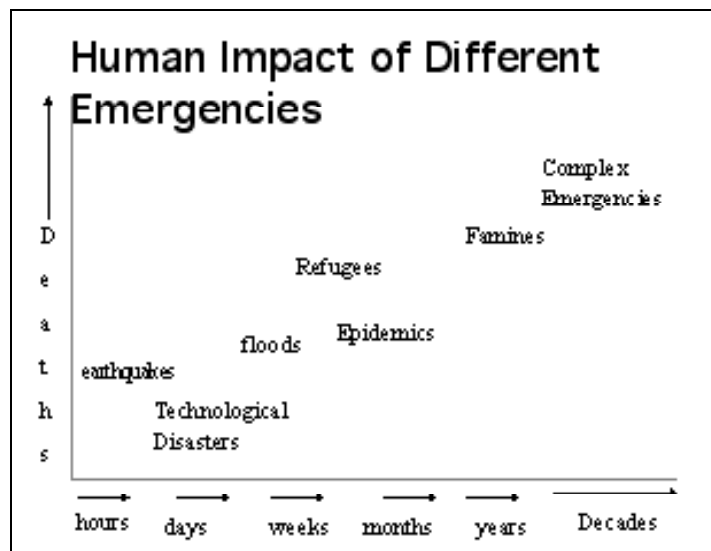


Figure 2: Human impact of different emergencies ranging from relatively short-term to long-term disasters.

Short-onset natural disasters – Floods, Earthquakes, hurricanes, tidal waves — lead to death from crush and suffocation, typically all within 50-70 hours of the event (Figures 2 and 3). The main program priorities are search and rescue, complemented by shelter from cold.

Most of the aid resources raised and delivered have nothing to do with saving lives: most of the deaths occur in the first hours only. Instead, most of the work of NGOs then is oriented toward long-term recovery, designed to rebuild in a manner to prevent future vulnerability to the same hazards. For example, prevention through more modern building design (using education, incentives), or in the case of Bangladesh, we have saved hundreds of thousands of lives through building of above-tidal-level structures in each village combined with better early warning of approaching tropical storms.

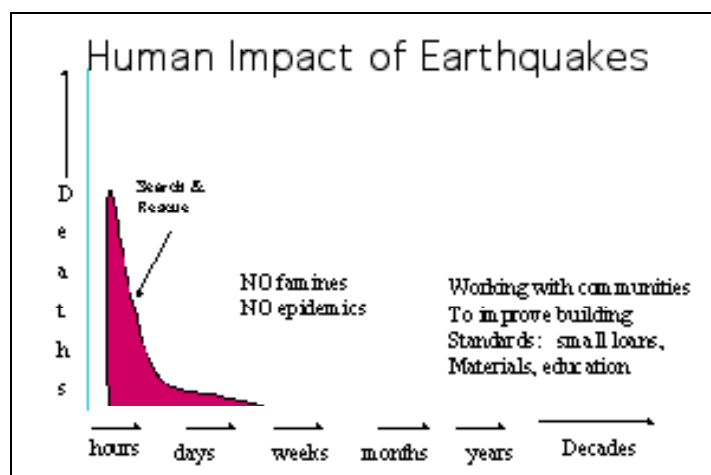


Figure 3: The human impact of earthquakes.

Famines and complex emergencies account for tens of millions of excess deaths in emergencies, in contrast to natural disasters which, in an average year, account for tens of thousands of deaths, several orders of magnitude less.

Famines and refugee movements play out over a longer-term. Drought stress lead to famine deaths only after the famine has evolved for roughly two years. Risk of death in a famine increases dramatically at the point where populations divest themselves of their property and move far as a desperation measure. Internally displaced persons, refugees, conflict victims, natural disaster victims, and environmental refugees all exhibit the same epidemiologic fact: being displaced increases death rate. The concept and term of “complex emergency” was coined in the late 1980s and, while not having a universal definition, such emergencies are characterized as having a combination of traits:

- High excess mortality.
- Multiple (over-lapping) threats or stressors.
- Violent conflict.
- Mass forced migration.
- Usually Access problems.

In Sudan, Angola, Mozambique, Somalia, Haiti, Tajikistan, Cambodia, Burma: it is the interplay of ages-old poverty with resource-competition-conflict with market failures and

natural disasters that kill. From the point of view of operations planning, the key aspect of Complex Emergencies that NGOs confront is their intrinsic unpredictability.

LIMITS TO OUR MEASUREMENT TECHNIQUES

We do not measure large country-wide crises well. We do not have sufficient techniques for understanding the scale and patterns of vulnerability or of outcomes (e.g. deaths) in large areas, such as Liberia, North Korea, Zaire, Sudan (Figure 4). We end up making absurd extrapolations that are, in retrospect, wildly inaccurate.

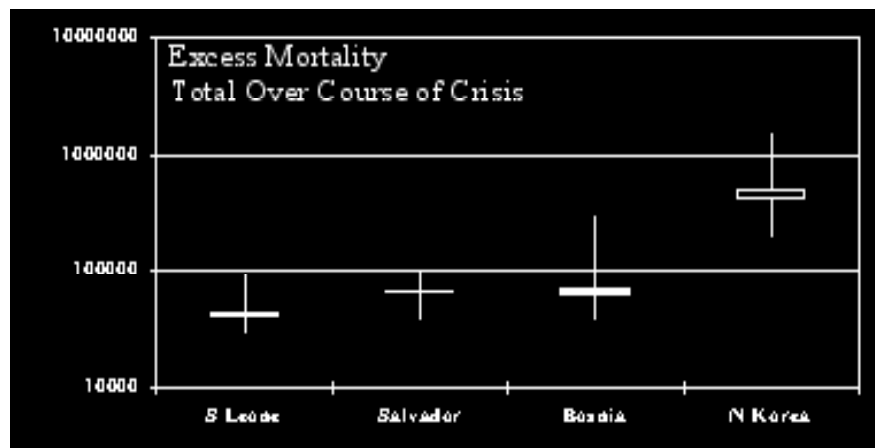


Figure 4: Variance in our knowledge of excess mortality exceeds our best estimates.

A subset of this challenge is that our assessment tools are ineffective at giving us information at a distance (e.g. cross-border) where we know how to view its accuracy or representative-ness. Cambodia in 1980 represents an excellent case; it was the celebrated, high-profile emergency that many agencies brought aid to, yet, in retrospect no one knows whether there was a famine in the first place, let alone whether lives were saved.

Whenever we evaluate our evaluations, we discover that NGO data on emergencies is patchy, it's better at looking at populations directly under our nose, much worse at understanding villages off the road, remote or outside of our aid targets.

There is no scientific or substantive basis for the "conventional wisdom" estimates of mortality in many emergencies. And where we have better data, there are huge disagreements among analysts.

We almost never make calculations of impact in our evaluations. So, for example, there is no basis anywhere for the figure of 70,000 to 80,000 deaths attributable to the crisis in El Salvador in the 80s. In Bosnia and North Korea, reasonable experts differ in their estimates of total deaths by many fold, even orders of magnitude.

Much of our historical estimates of total deaths from emergencies has its origins in estimates made up by journalists. And whenever two journalists have different estimates, the higher one always wins out in the print record. For this and similar reasons, therefore, there will always be an upward bias in what is documented and circulated of "numbers affected."

Worse, media will always find someone in every crisis to express “fear of famine and epidemics”, even though we know that in most natural onset disasters there will, predictably, be no epidemics or famine.

NGOs rarely calculate the impact (outcome benefits) of interventions – how many lives saved, in part because there’s no agreed-upon science or meta-analysis methodologies. In addition, we do not run control groups; we have no (or little) experimental design available, and thus unable to compare intervention groups against non-intervention groups.

CONCEPTUAL CONUNDRUMS

This is not a Magritte painting; this is a copy of an electronic representation of the Magritte painting (Figure 5). Magritte’s point (picked up on happily by philosophers) is that we frequently confuse our image or reference to a thing with the thing itself.



Figure 5: A copy of an electronic representation of a painting by Magritte.

Similarly NGOs frequently get tripped up through repeated use of an index that they confuse the index or metric with the thing itself. One example close to my hear is that we so often use weight/height as a clue to an individual’s malnutrition, that many nutritionists become limited by thinking that weight is malnutrition; in fact malnutrition is a multiple internal physiologic, organ and function failure. As another, similar example, our model for why kwashiorkor (a type of malnutrition) occurs is that it results from insufficient protein in the diet. But the simplistic response — to give the child a lot of protein in their recovery diet, which most physicians do, is lethal.

High protein diets are the worst possible thing for a severely malnourished child. Another example: we frequently speak of “food program” meaning that we are moving food. But just because food aid is an input to a project does not mean that the action or output of the project is “feeding;” frequently we sell food, or use it in food-for-work or subsidies or other market interventions. Food is fungible.

There is enormous bias in our data collection, our analysis and our reporting that NGOs have not successfully overcome. Among these biases:

1. Spatial bias
2. Project bias
3. Dry-season bias
4. Diplomatic bias
5. Professional bias
6. Mandate bias
7. Institutional bias

Chief among our operational biases is over-reporting of self-selected populations. For example, we frequently report the “level of malnutrition” in a famine feeding site, as if the information said something representative or even meaningful about a population. Yet these programs stipulate malnutrition as a “criteria of entry” so it is tautological, then, to report that, lo and behold, we discovered malnourished children in our center for malnourished children. This reporting confusion is not rare, it makes up some 50% of all the data flowing out of emergencies.

In reporting out, NGOs and UN agencies rarely mention the existence of other groups. Even though the same packet of aid (for example, a bag of food), will pass through the hands and responsibilities of a dozen different agencies, each one will claim sole responsibility for it as “their program.” Moreover NGOs are certainly aware that most of their work is done by local hires. And that most life-saving work is accomplished by the affected population itself, not by international aid interveners.

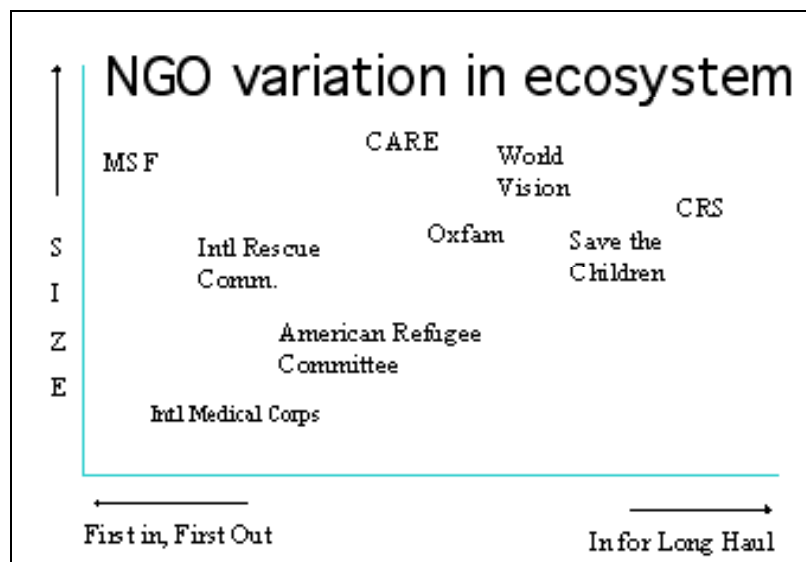


Figure 6: Non-governmental organizations become involved with different levels of effort and are active for different lengths of time.

NGO evaluations are not very sophisticated at recognizing the rest of the ecosystem (Figure 6). From one point of view, NGOs fill separate, distinct and important skill/task niches in a multi-dimensional space of what needs to be done to minimize mortality in an emergency. The system works because they don't coordinate (or standardize) too much. But if NGOs did better examination of what are the gaps — things left undone by the collectivity of NGOs — they would put more effort in to water supply and sanitation, the two problems that account for many easily-prevented deaths.

NEW AND EMERGING TRENDS

NGOs have been slower than the military to recognize and adapt available technologies, but are slowly learning to use GPS for location, GIS for vulnerability mapping, and Palm Pilots for rapid and accurate field data entry. Similarly, NGOs are gradually becoming more sophisticated about how to effect markets, to work with market systems instead of always treating the commercial sector as an enemy. For example, NGOs use food for work (old), monetization (newer), local purchase, and triangular transactions to move food around to meet food security needs. Reducing the price of food, in a famine, in and of itself, is a valuable outcome.

Finally, NGOs and their donors are increasingly adopting a “Rights Based Approach” to humanitarian interventions, led by the SPHERE project, a multi-year effort by NGOs from around the world to establish that disaster assistance is a fundamental, universal right of the people involved, not simply a technical “planning figure” that results from our dry assessments.