

Measuring Program IMPACT

LESSONS IN EVALUATION FROM HUMANITARIAN AID

Because political change does not come in units of length or weight, democracy promoters find evaluation an ongoing challenge. As in the field of humanitarian aid, the best solution might be an educated guess.

by Steven Hansch

Within the broad field of international aid, the sub-field of emergency assistance may well have established the most successful policies and practice for monitoring program outputs. Ironically, much of this progress has occurred despite a pervasive and ongoing ethic among emergency aid workers that the urgency of response should be unencumbered by ivory-tower academic studies. In addition, the moral environment of emergencies tends to dissuade responders from considering experimental designs, control groups or informed consent.

Given these circumstances, part of the reason for progress rests on humanitarian assistance's good fortune to include two disciplines that, on their own, emphasize measurement: nutrition and medical care. Nutritionists began documenting patterns and trends while assessing famines and refugee needs in the 1960s, after U.S. food and nutrition aid became a part of U.S. foreign assistance first under Herbert Hoover during WWI and then in WWII. Nutritionists reached agreement decades ago on the consistent use of measures

Aid agencies have been conditioned through years of reporting to donors, such as USAID, by accounting standards that focus on what is known, not estimated. As a result, aid agencies avoid reporting impact altogether and instead report softer data...

such as weight-for-height to allow precise statistical measures of a population. Repeated over time, these allow famine responders to gauge how well a relief effort is controlling problems of food insecurity.

In the 1980s, medical doctors, who had studied the emerging science of public health, began to apply epidemiological

tools and norms to the health aspects of emergency response, influencing first the Centers for Disease Control, then USAID and the United Nations, and eventually NGOs, to base their programming on the analysis of diseases and health outcomes. Increasingly, the application of statistical tools (promoted by epidemiologists) has led to more creative use of sample surveys. In failed states, refugee camps and other complex emergencies (where the true population size is unknown and unknowable), the solution for measuring changes over time rests in the liberal use of random-sample methodologies. Indeed, a random sample survey can provide a better estimate of the rate of a factor in a population than a comprehensive house-to-house census, because the random-sample frame can better adjust for and minimize key types of bias.

Measurement of health outcomes such as total deaths (or the crude mortality rate) lends itself as a useful index because it reflects, at once, the sum of failures in meeting basic human needs in all sectors. The rate of mortality is unambiguously defined and free of bias due to culture, response agency, or changing definitions over time.

Thus, aid agencies routinely gauge the scope and depth of an emergency in terms of indicators such as crude mortality, which can be tracked and compared from day to day. Weighting priorities and monitoring programs in humanitarian aid has been well codified in recent years, represented in the SPHERE humanitarian standards, which are signed on to by al-

most all NGOs, U.N. agencies and donors (such as USAID).

Yet, humanitarian aid agencies suffer a sort of writer's block when it comes to using these same tools to estimate what they achieved, i.e., their impact. A recent report from the U.K. Overseas Development Institute¹ finds that, while there has been increasing pressure on humanitarian aid groups to calculate what their ultimate results or achievements are, implementing agencies do not currently have the skill-sets, policies or practice in place to estimate levels of impact.

A key impediment has been the reluctance of aid agencies to infer. Even using hard data for inference or extrapolation, estimating against a counterfactual (e.g., how many people would have died had aid not been delivered), goes against past habit. Aid agencies have been conditioned through years of reporting to donors, such as USAID, by accounting standards that focus on what is known, not estimated.

As a result, aid agencies avoid reporting impact altogether and instead report softer data, such as how many people were helped, or output data, such as how many commodities were transported and delivered.

As a result of this avoidance of impact estimates, the community of emergency assistance agencies has built itself a tar-pit, where it's stuck—without any body of literature that offers any consensus about what "impact" even means from one agency to the next. Simply estimating the total number of deaths averted, say in a famine, does not seem sufficient or satisfactory for many emergency professionals, whose goals also include the reduction of suffering and protection of dignity. No aid agency has put forward a concept of how suffering or dignity might be scaled, let alone measured. Other impacts, such

as restoration of livelihood, reduction in the likelihood of war (conflict mitigation, or reconciliation) and return of displaced populations also are written about extensively as problems and goals, but without a proposed metric or scale.

In this frustration over a lack of suitable impact or results measures, humanitarian aid agencies share common cause with human rights organizations and democratization organizations, particularly as humanitarian aid actors increase their ventures into programming in rule-of-law, civil society promotion, transparency and peace-building.

At USAID, the combining of humanitarian, transition and democracy assistance into the same bureau (DCHA) four years ago lent more weight to the proposition that the solution to failing and failed states involves a synergy between life-saving assistance and improvements in governance.

Democracy organizations might consider borrowing some lessons from the emergency relief community—for example, the use of two-stage cluster samples in a large population to measure a commonly agreed-upon index. The emergency community finds common currency in the use of the “rate of malnutrition” (weight-for-height), where the data comes from a few hundred individuals sampled by a

two-stage cluster sampling method. In the democratization community, a common index might similarly gauge, in societies coming out of protracted emergency, “accurate knowledge among heads of households about how to register and vote.”

Similarly, emergency agencies perform “end use monitoring,” where a small sample of households are observed after the distribution of a monthly food ration to see how many received the full ration to which they were entitled. Democratization groups already use sample surveys for “exit-polling” after votes.

But both communities need to create a body of literature that estimates impact through application of meta-analysis, applying coefficients derived from global patterns to judge the impact of an intervention in individual instances. For example, in the health arena, we know from global experimental studies that the annual provision of vitamin A pills to children results statistically in the avoidance of something like one-quarter of deaths on a large population scale. Based on these facts, a relief agency can infer in many settings that if it distributes vitamin A pills to a large population of children, the resulting number of deaths will differ from the counterfactual (avoided) rate of deaths by about 20-30%.

Similarly, if it has been observed that a

package of democracy aid is followed by a certain outcome in long-term studies in 50 countries, then one can infer that a properly delivered package of aid in a new instance may have comparable impact, adjusted for what is observable.

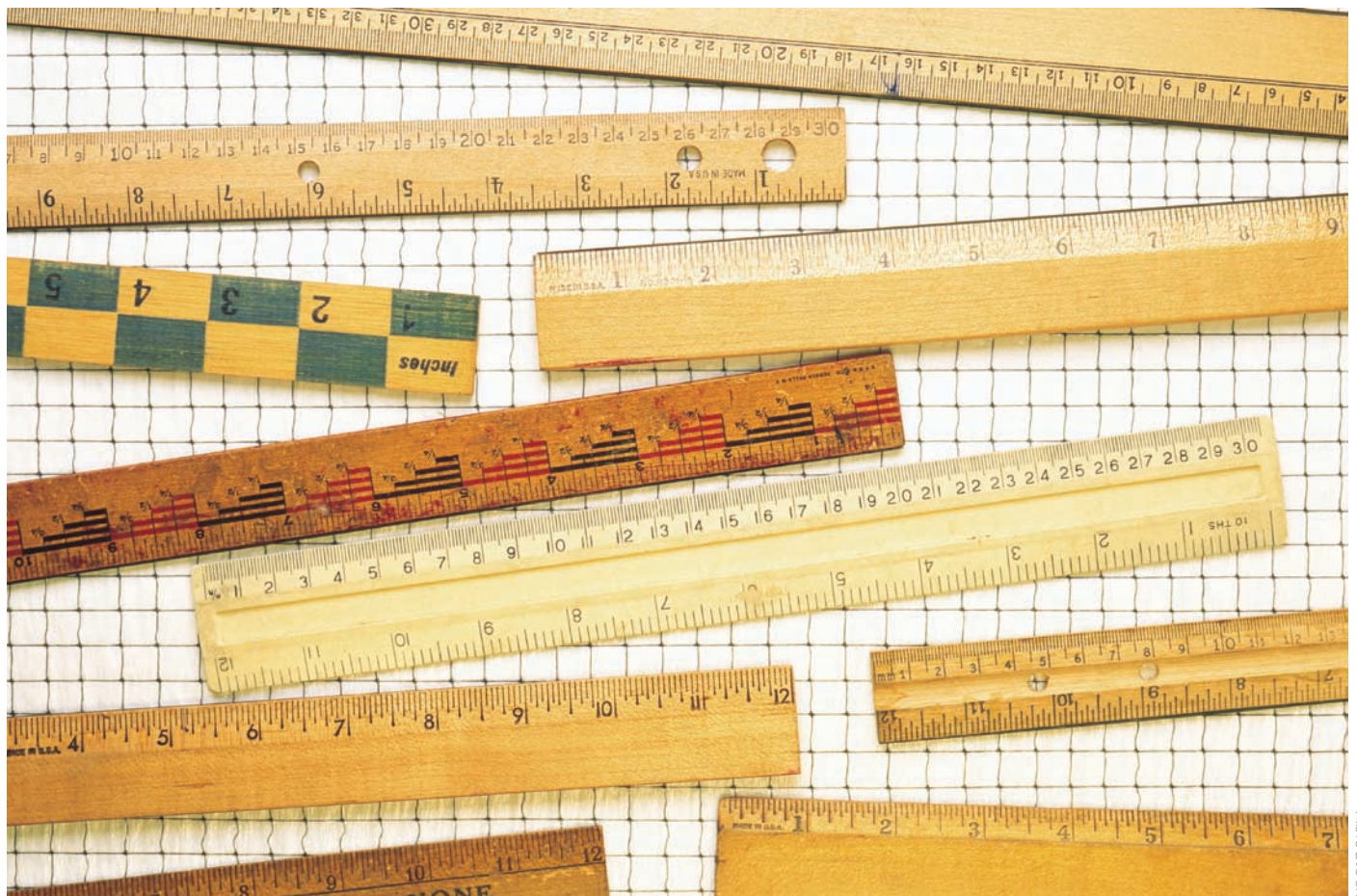
The solution to overcoming the hump in estimating impact may rest with jump-starting the debate through better collaboration with independent research groups, including universities. Sometimes it takes an institute that is impartial, and not operational in the delivery of aid, to propose its own scales (for example, for corruption or economic freedom), which can get the ball rolling for others.

A solution to this dilemma is critical to further maturation of the discipline of aid. A solution will allow a full accounting to the public, the ultimate donors, of what their return on investment was in foreign aid. And it will also allow a more rigorous and scientific determination of what works and what doesn't. [Read](#)

Dr. Steven Hansch is a Senior Associate at Georgetown's Institute for the Study of International Migration.

Notes

- 1 Charles-Antoine Hofmann, “Measuring the Impact of Humanitarian Aid” Humanitarian Policy Group Research Paper #15, (London: Overseas Development Institute, 2004).



AbelStock