Aldo Benini

A computer simulation of needs assessments in disasters

The impact of sample size, logistical difficulty and measurement error

November 2012

Suggested citation:

Benini, Aldo (2012). *"A computer simulation of needs assessments in disasters. The impact of sample size, logistical difficulty and measurement error"*. Washington DC.

Contents

Tables and Figures

Acronyms and abbreviations

ACAPS	Assessment Capacities Project
GIS	Geographic Information System
IASC	Inter-Agency Standing Committee
MIRA	Multi-Cluster/Sector Initial and Rapid Assessments
NATF	Needs Assessment Task Force
ODI	Overseas Development Institute
STATA	Stata - A statistical software
UNOCHA	United Nations Office for the Coordination of Humanitarian Affairs
VBA	Visual Basic for Applications

Acknowledgment

The computer simulation and a large part of this paper were originally part of a larger project that I conducted for ACAPS in 2011. I thank Patrice Chataigner, ACAPS, for permission to use elements of the project report here. However, all errors, factual and of interpretation, are mine; no responsibility for ACAPS is implied.

The maps in Figure 3 were created by Olivier Cottray, then with iMMAP in Geneva, Switzerland. Christian Cricboom, then with UNOCHA in Islamabad, Pakistan, shared data of an assessment of flood-affected Union Council areas in Punjab Province. I have used his rankings to assign communities with simulated impacts to a spatial structure. I am grateful for this assistance.

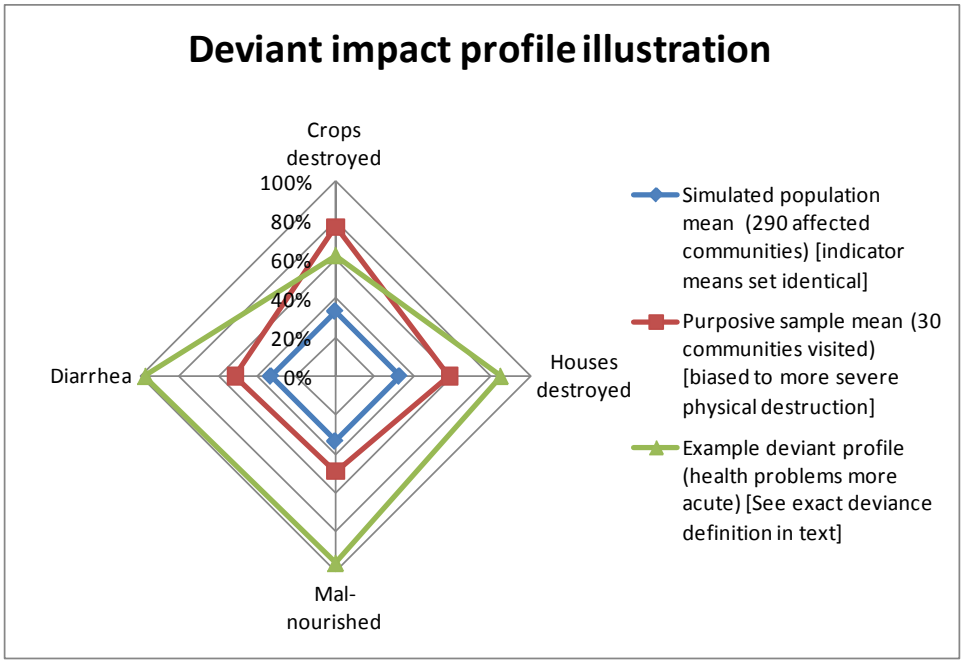
Aldo Benini

Summary

United Nations agencies, principally UNOCHA, and NGOs concerned with disaster management have moved towards a four-phase system of needs assessments after sudden onset-disasters. Objectives and methods change over a period lasting more than one month: *"Phase 1 - Saving lives", "Phase 2 - Sustaining lives", "Phase 3 - Stabilizing lives", and "Phase 4 - Restoring livelihoods"*. Phase 2, also known as "Multi-Cluster/Sector Initial and Rapid Assessment" (MIRA), takes place in the first two weeks after onset.

Phase-2 Primary Community Level Assessments were simulated in a spreadsheet environment. Assessment teams follow itineraries that optimize the rapport between prior impact estimates and distance to travel. The key variable of interest is the frequency of deviant (= atypical) community impact profiles that add insight into the diversity of situations among affected communities. The computer simulation was part of a larger study of data management and data analysis in Phase 1 and 2 assessments.

Figure 1: Example of a community impact profile



Twelve models were run, varying three parameters: sample size, logistical difficulty and level of measurement error. Each run produced statistics on fifty possible itineraries. Impact profiles were characterized by comparing the values on four indicators that the teams had estimated on site. An estimate consists of a simulated "real value" plus the measurement error. The "real values" had been generated using a correlation structure imposed on the indicators. The spatial structure was borrowed from a UNOCHA study of flood-affected communities in the Punjab Province, Pakistan.

Key findings from the comparison of model outcomes include:

- Chances to discover novel impact constellations increase with sample size, as expected. The marginal effect decreases after a certain point. Assessments can purchase higher qualitative complexity with more information, but this comes at an escalating cost.

- Surprisingly, more difficult logistics improves those chances. These difficulties compress itineraries, causing teams to focus on smaller clusters of communities instead of moving to farther places selected on already accustomed criteria. Within the smaller clusters, some communities with lesser values on dominant indicators will reveal higher values on indicators that initially were less prominent. This stimulates the recognition of novel impact constellations.

- Measurement error too produces more communities with deviant profiles. It does so by adding to the variance in the indicators. But this diversity is largely fake and is potentially harmful.

Practically, the results suggest trade-offs by spending more time in a given community (to reduce measurement error) and working in smaller clusters (less travel time), and only secondarily by increasing sample size.

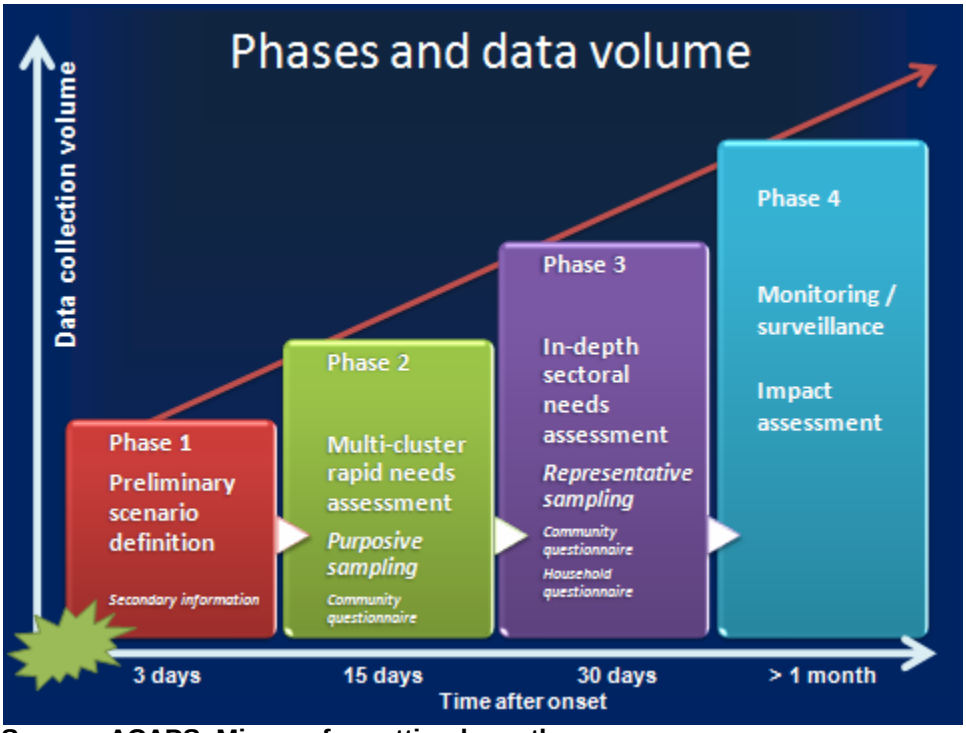
The realism of some of the simulation assumptions is questionable. Nevertheless, the effects of parameter variations on key outcomes appear worthy of discussion. Also, the simulation draws attention to measurement error, a concept central to survey research, but not one prominent in the needs assessment community. This or similar simulations may have a place in the training of assessment personnel.

Introduction

The humanitarian community has invested considerable effort in improving the assessment of needs among disaster-stricken populations. Within the United Nations, UNOCHA leads the development of methodologies for what is known as *"coordinated multi-sectoral needs assessments"* (NATF 2010: 2).

These shared efforts to acquire situational knowledge about disaster areas, especially in the wake of sudden onset disasters, place a strong premium on speed, data exchange and sufficiently complete analysis. A consensus has emerged towards a phased approach. The model of four phases - or rather five, including pre-disaster preparedness - has been described in various documents and has lately received added prominence in a dedicated Overseas Development Institute publication (Garfield, Blake et al. 2011). The naming of the phases varies - the ODI study, for example, characterized them by their paramount goals: *"Phase 1 - Saving lives"*, *"Phase 2 - Sustaining lives"*, *"Phase 3 - Stabilizing lives"*, and *"Phase 4 - Restoring livelihoods"*. Others prefer terms that have informational connotations, such as *"Preliminary scenario definition"* for Phase 1 and *"Initial Rapid Assessment"* for Phase 2 (NATF 2010). More importantly, the timing of the four phases in large degree dictates the options available for the organization and contents of needs assessments. These products are considered useful only if they are shared within very tight timelines - the preliminary scenario report within three days from the disaster onset, the initial assessment report within two weeks, an expanded assessment within 45 days.

Figure 2: Assessment phases



Source: ACAPS. Minor reformatting by author.

Assessments as organized learning

Needs assessments are organized learning processes. The nature of the disaster and the concerns of the responder community define a focus. The focus makes perception, analysis and reporting highly selective. At the same time, because the diversity of disaster impacts and of survivor needs is not sufficiently known, assessments should remain open for the unexpected. This openness, however, competes with timelines and with the discipline of applying pre-defined cognitive and administrative categories. Novel information, if registered at all, demands added effort to record, structure and evaluate. If recognized midway through an assessment phase, the novelty may call for confirmation in additional sampling points.

Assessment planners and, in the course of execution, field teams must make compromises. There is little guidance available from the existing literature. For illustration, we look into the challenge of purposive sampling, a method prescribed for the selection of affected communities to visit during Phase 2. Depending on which communities they choose to visit, teams will chiefly confirm the initial assumptions that guided the sampling, or identify deviant impact profiles modifying those assumptions.

We use computer simulation as a technique to explore the ability to detect less common needs profiles. We build a model of an assessment team deciding a circuit of sites to visit. The team selects subsequent sites considering partial knowledge of the degree of impact as well as the effort to travel there. We vary a small number of parameters to estimate the probability that the team will visit communities with other humanitarian needs about which it knows nothing beforehand. Cognitive processes - what the team perceives and reports - are difficult to model. Decision processes using incomplete information, however, are more easily accessible to simulation.

We proceed as follows. The next section recalls specifics of Phase 2 assessments. We then describe assumptions and set-up of the simulation. We explain the parameters varied across runs of the simulated itineraries that the assessment teams follow across affected communities. In the findings, we summarize mean effort (distance between visited communities) and the number of communities with atypical impact profiles that the teams met on their circuits, in response to sample size, logistical difficulty, and measurement error. We discuss the significance of the findings. We conclude by reminding readers that simulation assumptions, including ours, are questionable. We draw attention to measurement error as a neglected concern in the humanitarian assessment community and at the potential use of such simulations in assessment trainings. An appendix details some of the technical considerations; others are spelt out in supplementary material.

Assessment concerns in Phase-2

Phase 1 of the assessment helps define a "preliminary scenario" of the disaster, within three days of the onset. Phase 2 aims at an "Initial Rapid Assessment" to be developed over the first fifteen days.

The assessments in Phase 1 and Phase 2 not only have different objectives, they also face fast-changing information environments. In Phase 1, high uncertainty pervades our

knowledge of the disaster, and the dimensionality of the core information is low - primarily estimates of affected populations by region, vulnerable group and needs sector. The sources of information are typically secondary.

In Phase 2, scale, scope and space appear more clearly. The information volume is higher. Phase 2 takes place in the first two weeks after disaster strikes. Typically, there is now enough access to, and communication from, affected groups to shift the balance between primary and secondary information. The focus turns to assessments of *"needs, risks and priorities as perceived by affected populations and assessment teams"* who visit a number of local sites (NATF 2010: 6-7).

Goals and objectives too change. At the goal level, the "common multi-cluster initial assessment" is to establish humanitarian priorities across sectors (UNOCHA 2009: 17), inform the response planning and orient follow-on assessments. The focus is still on disaster-wide priorities, not on the needs of individual communities, despite occasional misunderstandings about the function (see, e.g., on Haiti: Julmy 2011).

The objective is to produce an integrated view of needs profiles from community-level input, the opinions of assessment teams who visit them and those of other key informants, as well as from the continued inflow of secondary information. The dimensionality of the core information is now much higher than in Phase 1. It grows roughly in proportion with the length of the questionnaire that bundles humanitarian indicators and other items of interest. This complexity may be further inflated with variables that hold assessor opinions, pre-disaster indicator levels (if they are elicited), and/or concurrent gender-specific focus group ratings. In addition, units of measurement may grow heterogeneous. While uniform ones are sought - communities or institutions (NATF 2011: 18) -, some key informants and much of the secondary information will speak to conditions of a whole district, region or social group. Communities may be split up by displacement, with old-site and camp groups, and with camps filled from multiple villages and towns.

During their visits to affected communities, the assessment teams elicit statements on impacts and corresponding unmet needs across several sectors. This basic set-up forces choices on two sides of data production - the instrument with which the teams measure needs, and the selection of sites at which they measure needs.

In this paper we are not concerned with specific techniques (see Benini 2011), but with measurement quality in general. Contrary to the survey profession, for which error control is a key concern (Groves 2004), the humanitarian assessment culture is not highly error-conscious. The design of humanitarian indicators is motivated by substantive considerations, and much less by classical measurement issues. Yet, these - whether validity issues between construct and measurement, measurement error between stimulus and response, or processing error between response and edited data - do affect the quality of the assessments. A general awareness of measurement error will be helpful and will probably by itself put a brake on the complexity of assessment designs.

As regards, the selection of sites, the current (and foreseeable future) Phase-2 assessment methodology calls for visits to a *purposive* sample of affected communities. Such a sample *"cannot be generalized statistically, but can nonetheless provide critical information on areas and groups in greater or lesser need"* (Garfield et al., op.cit., 13).

Guidance on purposive sampling is not very detailed. Meaningful stratification is advised *"to ensure diversity and systematic comparisons among relevant groups"* (ibd.: 15). Sample size, one assumes, must ultimately depend on logistics, staffing and time.

Detecting diverse needs

Phase 2 *"is a period where it becomes more and more clear that different areas and populations are differently affected",* as the Working Group recognized. Thus, assessment teams require an appropriately complex understanding of the diversity of impacts. The ultimately mapped diversity springs from three sources: team members' conceptual backgrounds; Phase-1 findings and the language in which these are expressed; the patterns recognized in the course of data collection in the field and subsequent analysis.

Specifically, we simulated the ability of Phase-2 Primary Community Level Assessments to produce informative impact profiles of disaster-affected communities. "Informative profile" in this context means community descriptions that qualitatively highlight distinct constellations of different impacts across a sample of communities. By contrast, "informative" does not mean "representative". Teams are drawn to more highly impacted communities and thus return purposive samples that cannot be generalized to the population of affected communities.

The simulation

Set-up

For this simulation, we simplify the set-up both analytically and empirically. We express impact at the community level by proportions of particular universes affected by particular disaster effects. Examples include the percentage of residents taken ill or the percentage of cropland destroyed. These are observable indicators; their amounts are thought to be produced by underlying factors such as physical force or stressors on the human organism. While the underlying factors are not fully observable (or not yet at the time of the Phase-2 assessment), they have correlated effects open to assessment during this period.

The combination of values that an affected community displays on impact indicators can be called an impact profile. A particular type of disaster may cause a majority of the affected communities to respond with profiles that outside observers recognize as typical. Floods, for example, destroy buildings and crops; these may initially be perceived as the dominant impacts. Across the range of affected communities, however, some may exhibit profiles deviating from the typical one. If many of them display deviant, yet similar indicator values, chances are that this group will be recognized for a distinct impact profile. The diagram in the summary section (page 6) illustrates this with an example

from a simulated assessment circuit; the green lines stand for a community with atypically acute health problems.

Before assessment teams move in the field, assumptions are formed in the course of the Phase-1 scenario formulation. These and site-specific information such as from satellite imagery provide initial impact estimates. In the field, the teams follow itineraries that optimize the rapport between prior impact estimates and the effort it takes for local measurement. In the simplified perspective of this simulation, effort is expressed as a function of the distance to travel to the affected sites.

The structure of impacts too must be simplified for the purpose of simulation. But it cannot be reduced to a single variable. It is reasonable to assume that all major impact types in a homogenous[1] disaster are fairly strongly correlated. In other words, many, but not all, of the communities highly impacted in dimension A will show high impacts in dimension B as well. Some, however, will deviate from this pattern. If several of the deviants have similar impact profiles, they will likely be recognized as a typical, if minority, profile in their own right. The key variable of interest is thus the frequency of deviant (= atypical) community impact profiles that add insight into the diversity of situations among affected communities.

Correlation structure

In simulation terms, a core component of the set-up is a suitable correlation structure for the observable impact indicators. For a minimalistic argument, we choose four: two more strongly correlated with one causative factor, and two with a different factor, and a weaker correlation between the two groups.

For illustration, our four indicators are from a flood disaster setting. It is important to point out that this is only one of any number of potential substantive interpretations. Thus, the realism, or not, of the model has little to do with the peculiarities of floods. The same correlation structure could be imposed, if one so wished, on sets of indicators pertinent to other types of disasters. Figure 1 shows the 2-factor 4-indicator set-up with the associate correlation matrix.

[1] The Pakistan 2010 flood exemplifies a homogenous disaster, as different from, say, Japan in 2011, with its combination of earthquake, tsunami and nuclear accident.

Table 1: Impacts, indicators and indicator correlations

Basic impact-producing mechanism	Observables (indicators)	Simulated random variables	Crops destroyed x1	Houses destroyed x2	Children under-nourished x3	Persons with diarrhea x4
Physical destruction by flood	Crops destroyed	x1	1	0.8	0.5	0.5
Physical destruction by flood	Houses destroyed	x2	0.8	1	0.5	0.5
Stress on human organism	Children undernourished	x3	0.5	0.5	1	0.8
Stress on human organism	Persons with diarrhea	x4	0.5	0.5	0.8	1

Spatial clustering

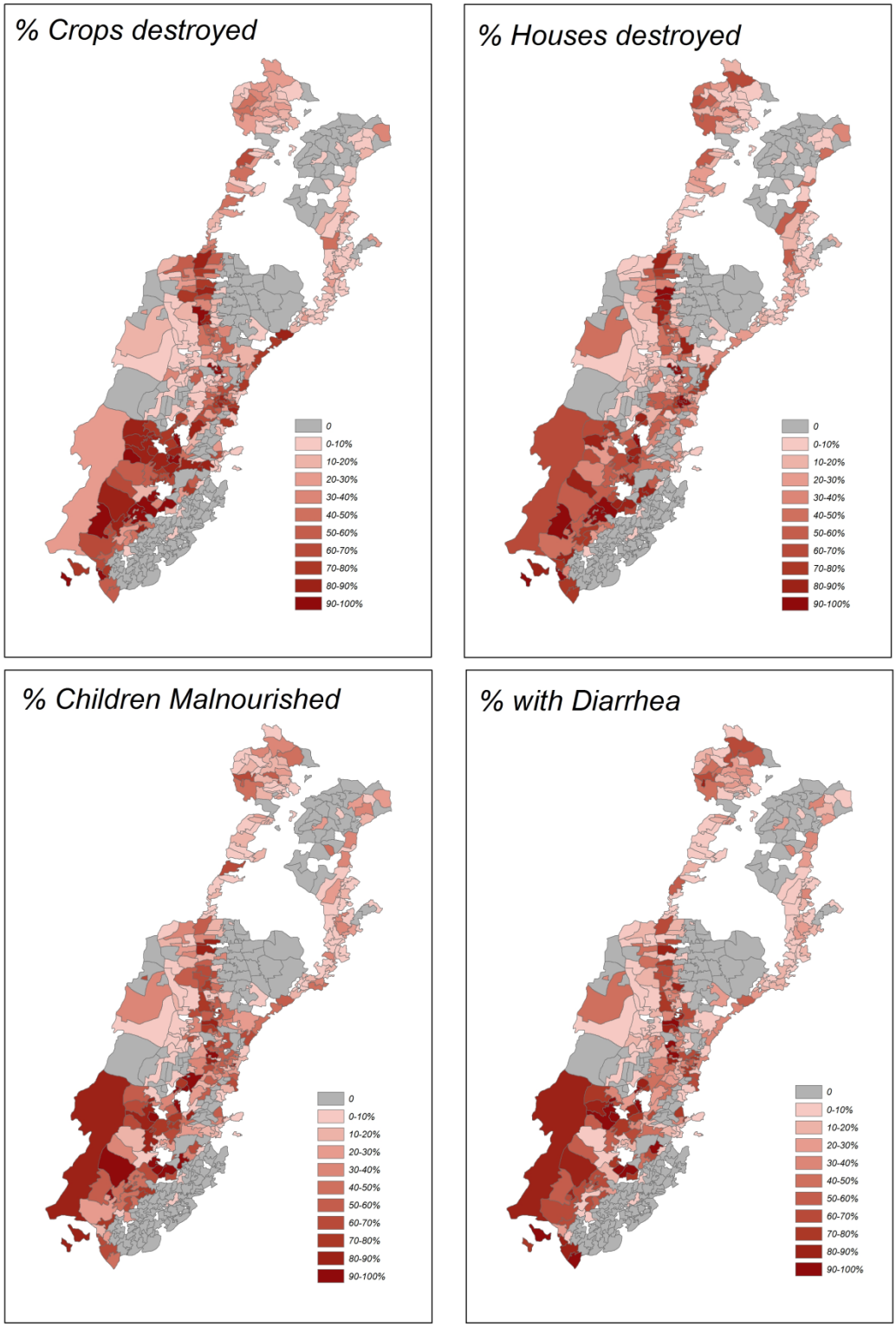
The second core element of this simulation is the spatial structure of the impacts. Disasters may have remote ultimate causes, but their intermediate factors and observable impacts tend to have a localized spatial structure. One could, theoretically, simulate such a structure on an abstract chessboard-type community grid. This would not be didactically attractive. Opportunistically, we make use of an impact assessment that the UNOCHA Pakistan team conducted, and subsequently mapped, of flood-affected communities in the Punjab Province, in December 2011 (UNOCHA 2011). Their methodology, working with district-based meetings of knowledgeable stakeholder representatives, produced an impact score for each of over 300 affected communities. We found spatial references for 290 communities.

The "Total Score", as constructed by the UNOCHA team, is only partially valid as an impact measure[2]. Nevertheless we assume that the spatial structure of this index is strongly correlated with some valid impact measure. For simplicity, we assume that the Total Score is monotonic in some function of our four impact indicators if we simulate variates for the same number of communities. In other words, there exists some impact index formed from our four indicators such that, if community A ranked higher on it than community B, the UNOCHA total score for A would be equal to, or greater than, for B.

Practically, we merge our table of simulated indicator values to the UNOCHA table by the descending rank of the mean of the four indicators, respectively of the Total Score. Our simulated data thus conveniently inherit the spatial structure of (part of) the Pakistan flood disaster. This provides a measure of didactic plausibility (simulation results can now be mapped out) as well as the geographic information needed to locate and fashion simulated assessment team itineraries. It says nothing about the validity of the simulation results (or, for that matter, of the UNOCHA measurements). We use the centroids of the 290 Union Council areas as destinations of the assessment teams.

[2] For reasons spelt out in a separate note to its principal author and to ACAPS. This is also why we do not use the information contained in the individual indicators collected by UNOCHA.

Figure 3: Spatial clustering of indicator levels among 290 affected communities in Punjab Province (simulated values)



Behavior of assessment teams

The third key element consists of the behavioral assumptions regarding the assessment teams. We work with one team only. We do not believe this entails loss of generality since the affected region could be partitioned, with each section assigned to one team, who would then follow the same rules.

We assume further that, prior to site visits, some impacts are easier to observe than others. Specifically we assume that levels of physical destruction are more easily observable - by means of fly-overs or key informants interviewed in administrative centers - than health impacts[3]. We assume that the assessment team selects, as a starting point of its itinerary, one of the communities reputed highly impacted. Once it has traveled there, it determines the next community to visit by weighing the estimated impacts of, and distances to, other affected communities. The artificiality of this assumption is blatant; in reality, teams travel on roads, waterways and airways still open.

We make one further assumption: That the assessment team has perfect information on the physical destruction levels of all affected communities, but has no information of the health situation whatsoever. The total lack of realism of this assumption needs to be noted upfront. We make it nevertheless because we wish to model the effect of measurement error at a different point in the process: once the team has arrived at the selected community and estimates indicator levels on the spot. This arrangement allows us to locate uncertainty in one particular point of the assessment process and to vary levels of uncertainty there. It also dispenses with the complexity of modeling learning effects that take place while the team travels (and which in reality, one hopes, will be significant in any team worth its salt!).

In the ultimate step of the set-up, we define a quantitative measure of the ability to differentiate impact profiles at the community level. Since the team selects communities entirely on observable physical destruction (and, if distances were identical, would always go to the next community that has suffered more severe destruction than others), naturally the question arises: What are the chances to meet communities that have, relatively speaking, higher health impacts than destruction levels? These chances are not insignificant because in selecting communities for the itinerary, physical destruction competes with distance. But, on what exactly do they depend?

Key outcome variable

While in the community, the team estimates impacts from physical destruction as well as in the health sector. We set an arbitrary level of difference between the mean of the two health indicators and the mean of the two destruction ones in order to define a deviant

[3] The situation has a feeble affinity to "principal-agent" concepts with multiple agents, some of whom have access to types of information than others don't, and vice versa (Pereira 2009). The assessment team, reporting to the United Nations and the humanitarian community, may be likened to one such agent; local authorities, chiefly accountable to domestic constituencies, are agents with different information.

impact profile[4]. In other words, if a visited community exhibits higher levels of health problems than destruction, it has a deviant impact profile.

The key outcome variable then becomes the number of visited communities having a deviant or, as some may prefer, "atypical" impact profile. One assumes that the team will formulate the existence of a qualitatively different impact profile - in our case, one dominated by health problems - with increasing likelihood as their assessment itinerary leads them to more such communities. If the team meets only one, the case may be filed away as a simple outlier. With two or more such discoveries, the team is more likely to recognize the distinctiveness of impact profiles. This assumption is not theoretically grounded here and would need to be based on more solid knowledge of organizational cognition (e.g., Meindl, Stubbart et al. 1996). We simply make the common-sense case that a novel category such as a distinctively different impact configuration is the more likely to be formulated the more instances appear that are in agreement with it.

Simulated itineraries

Our simulation thus produces many possible itineraries that an assessment team may follow. It computes and records certain key outcomes, and summarizes these for different parameter settings.

The key outcomes are:

- The average distance traveled between communities
- The number of communities with a deviant impact profile

The latter is computed from the estimates that the team makes of the levels of the four indicators while in the community.

The key parameters that are varied, with outcomes subsequently compared, are:

- The sample size (the number of communities visited)
- The level of measurement error in the on-site estimates
- The level of logistical difficulty

The difficulty of moving around from one community to the next is modeled with a power function of the distance, in analogy to common models of gravitational attraction or distance decay in social geography[5] . The level of measurement error is controlled

[4] Technically, this is feasible as long as all four indicators are drawn as random variates from the same distribution. A more realistic generative process would, at least, rescale some of the indicators. By reweighting the indicators, however, the same kinds of comparisons could be made. To keep things simple, we did not rescale, at the expense of plausibility. Nobody would accept, for example, that over 90 percent of the people in a community suffer from diarrhea at the same time. But this is a matter of technical convenience, not one that imperils the validity of the model.

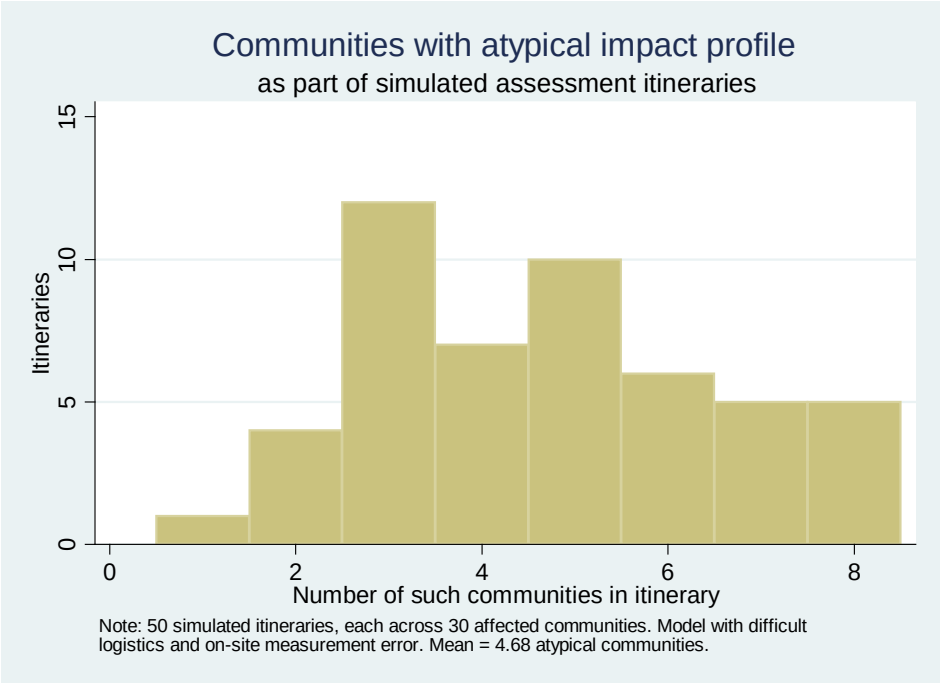
[5] Among others, Fotheringham et al. (2000: 225 sqq.) discuss spatial choice models and the conceptual difficulties associated with some of them. In reality, decision makers, such as our assessment team, cannot consider all alternatives (290 affected communities) simultaneously. The function of distance in the decision model then is to screen out irrelevant alternatives. That said, this simulation causes the team to

through the multiplier of noise variables added to the "real" impact variables; for this study only two settings were tried out: zero (no error) and normal variates with standard deviations half of those of the "real" impacts.

We also vary the starting point by randomly selecting one from the 50 communities remotely estimated most impacted.

Figure 4 illustrates the outcome of 50 simulated assessment circuits, each with 30 communities visited. This particular model assumes difficult logistics and on-site measurement error. The number of communities with an atypical profile among the thirty ranged from 1 to 8.

Figure 4: Variation in the key outcome across 50 simulated itineraries



The technical implementation is documented in part in the appendix, in part in comment lines in the table of contents sheet as well as the Visual Basic module attached to an Excel workbook. Here we report some key findings.

Findings

Twelve models - each defined by a combination of key parameter values - were run as simulations each of 50 possible itineraries. The distances that teams would cover between successive communities in their itineraries depends on logistical difficulty and on sample size. Measurement error has no influence on distance since it occurs on site only, after travel to a selected community. Larger samples, as expected, send up the mean distance

---

select the one next community at a time, and then another, etc., each time on the optimal rapport between the key informant estimate of the destruction and the distance from the community last visited. This can lead to zig-zag patterns. In other words, this is not an optimal routing approach.

traveled because with more communities already traveled, the pool of nearby highly impacted ones dwindles. As logistics eases, teams are more disposed to travel farther in order to meet high-impact communities.

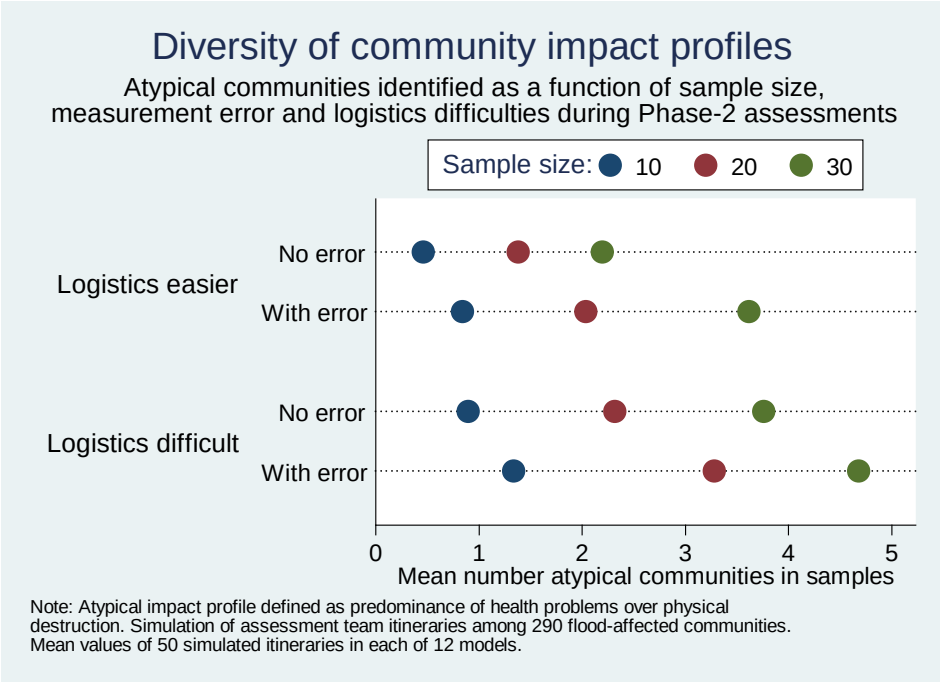
The results agree with those considerations. The exact figures, of course, are the outcome of the specific spatial clustering of impacts in Pakistan, which we borrowed from the UNOCHA data; they would be different in other settings even if our key parameters remained the same. As such, these distance findings are not of a great general interest. Nevertheless, note the small size of the differences, particularly as the samples become larger. This reflects the strong clustering of highly impacted communities, with very slow exhaustion of nearby ones not yet visited.

Table 2: Mean distance between visited communities (in km)

Sample size	Logistics	
	Easier	More difficult
10	12.9	10.5
20	13.4	10.9
30	13.9	11.1

Findings of more general interest concern the number of communities with significantly higher health problems (compared to physical destruction). We take this number as a predictor of the probability that the assessment team will formulate an additional distinct type of impacted community. These outcomes respond not only to logistics and sample size, but also to the errors that infect the measurements on site. We compare absolute counts rather than proportions of such communities because we assume that relatively few occurrences (> 1) are liable to get noticed as instances of a distinct category.

Figure 5: Communities with atypical impact profiles



Variable returns to sample size

Three findings strike the eye. As expected, as teams visit more communities, chances to find ones with deviant impact profiles improve. The marginal rates at first increase, then decrease for larger samples (with the exception of models for with measurement error and easier logistics, probably a random effect). This suggests that larger qualitative complexity can be purchased with more information, but at an escalating cost.

Smaller clusters enhance diversity

The true surprises arrive from the logistics and measurement error side. Estimates of deviant profiles become more frequent with difficult logistics and under error. The combination of both accelerates the rate considerably.

While few of us may anticipate those effects, they are plausible. Logistics troubles force teams to stay within smaller clusters. These exhaust choices of severely destroyed communities faster. Since smaller clusters tend to contain more communities with lesser physical destruction, chances improve to discover communities with more salient health problems.

Measurement error creates fictitious diversity

The mechanism for measurement error is different. Higher levels of error create perverse effects. Formally, they add to the variance in the observed indicators. Substantively, they produce fake diversity.

In other words, higher levels of measurement error may create the illusion of dealing with distinct types of impacted communities when these do not really exist, or only in such rare instances as not to warrant the explicit distinction. Or, distinguishing types may be justified (say, from a disaster response perspective), but a significant portion of the observed communities get misclassified due to on-site errors. One might argue that this matters little in Phase-2 assessments, which are to reflect the qualitative complexity of impacts. But significant misclassification will disguise the spatial clustering, if it exists, of communities with particular impact profiles, or will suggest such clustering where none really exists. And if this happens, it surely detracts from the assessment quality.

The significance of the findings

The simulation has given us three results as regards the detection of atypically impacted communities that may lead to the formulation of new qualitative impact statements. First, a minimum number of communities need to be assessed in order to find enough with deviant profiles to provoke additional categories. Larger samples make it likelier that this threshold will be crossed, but they also escalate the unit cost of discoveries (and, of course, meet with practical limitations).

Second, far-flung exploration of affected communities, maximizing selection by initial criteria (e.g., physical destruction vs. distance traveled), may not always be the most efficient route to new qualitative findings. Exploring smaller regions, compressed by greater logistical resistance, may seem a disadvantage. Yet, teams may thus have occasion to meet communities that are less impacted under the initial working definitions, but which exhibit other problems that enrich the understanding of disaster impacts. Subsequently, these may become part and parcel of the conceptual lenses through which the assessment and responder community look at the collectively of impacted communities.

Third, measurement error may lead to the discovery of patterns, but it may also be very costly. It may throw up new hypotheses (such as that communities of a new, distinct impact profile exist). It may as well attenuate existing and substantively important correlations so that they are no longer noticed (such as spatial clustering of certain problems). Errors may lead the assessment down the wrong path. They may be detected and corrected in later phases only, possibly after bad data misinformed response planning.

Trade-offs

In terms of trade-offs in organizing assessment itineraries, these findings suggest that savings might be most effectively made in the distances traveled. Visiting clusters of communities contained within smaller radii may uncover more genuine impact diversity than maximizing the enveloping region does. Travel time savings should primarily be turned into lessening measurement error by spending sufficient time in any one community being visited, and secondarily by increasing sample size.

Conclusion

This simulation mimics the assessment process in affected communities in Phase 2 of the MIRA. It was motivated by the question of how effectively the purposive sampling used in this period works for the detection of diversity of impacts, an objective stated sometimes in terms of "qualitative" research.

Questionable assumptions

In order to model the diversity of assessment findings, however, we had to give quantitative definitions of distinct impact profiles. We did this, by way of illustration, with a system of correlated indicators, two of which related to one particular mechanism (destruction by flood waters), and the other two to a different mechanism (stressors on the human organism). These indicators were used in the UNOCHA Punjab Union Council assessment. Our simulation borrows the spatial clustering from this study. Regrettably, because the correlation structure of its indicators is not usable, we replaced it with an arbitrary one that holds analytic interest. This is clearly an inferior make-shift approach. A plausible correlation structure among disaster effects still seems a good starting point for an assessment simulation.

Our simulation makes assumptions on how assessment teams behave. These are equally problematic. "Partial observability" - teams can assess the physical destruction, but not the health problems, before they meet the communities - has a robust literature in econometrics and decision theory, but this does not mean that our rigid assumption is correct. At most, it is convenient in simplifying the model. Similarly, the ways teams select the next community to visit is highly unrealistic. It does privilege highly affected communities, but at the expense of practical routing.

The most serious objection, however, concerns the way assessment teams learn. The implication of our model is that learning - the formulation of additional distinct impact profiles - takes place at the end of the itinerary, after teams count frequencies of different impact combinations and decide that a novel one occurs significantly more often than mere chance would have it. This is a very impoverished and probably also impractical view of human learning. Much more likely, teams learn in rather haphazard jumps and leaps during the assessments, whenever drastic impressions or stark contrasts with expectations force conceptual revisions. The only justification for upholding this assumption is that if communities with a deviant profile are visited in greater number, they provide more opportunities for such revisions.

Attention to measurement error

A strength that some may recognize in our approach is its incorporation of measurement error. Although the rapid assessment community claims to be doing surveys - starting in this Phase 2 -, its documentation is virtually devoid of all references to this key concept of survey research. It is as though the qualitative perspective on Phase-2 information collection miraculously did away with the possibility of error. In reality, of course, qualitative statements can be as incorrect as quantitative ones - by degree or in total. For lack of time, we have modeled the measurement error in a quantitative way. It would

deserve to be explored with other methods as well. Also, we ran models only either with no error or with one given level of error. Given more time, investigating models with at least three different levels of error would be more revealing.

Limitations and potential

More fundamental criticisms can be leveled at this simulation approach. Notably, the correlation structure among the disaster effects - here captured as "indicators" - must be the result of something. There should be a data generation process that parallels the known disaster causation processes. Taking the impact indicators as the point of departure short-changes this dynamic. This defect shows up in the fact that the random variates of the indicators are generated using their correlation structure, but the spatial clustering is entirely borrowed from an outside instance (the Pakistan floods) with very tenuous analytical connect (the only thing we can assume is that if our generative process took place on the Punjab community set, the UNOCHA impact measure would be monotonic in some function of our indicators). A more ambitious approach would be to generate impact correlations and spatial structure in an model integrated from the start.

Does this simulation approach have potential for other useful explorations? At a minimum, some of the code may be recyclable for the benefit of other simulationists daring into the needs assessment jungle. Second, some of the limitations of the purposive sampling strategy might be explored using similar simulations. This strategy, and the implication that no inference can be validly made to the affected population (Garfield, Blake et al. 2011), almost has the status of a religious dogma. However, a combination of simulation and adaptive sampling techniques (Thompson and Seber 1996) might release this constraint, if only for estimates of modest precision. Similarly, routing algorithms, some of which are built in with GIS software, could be used in simulating efficient assessment itineraries. To move towards any of these frontiers, however, it will be instructive to take stock of what Phase-2 assessments have achieved in recent disasters, and how they did it.

In addition, the simulation model, since it is spreadsheet-based, could be adapted for training purposes. Assessment personnel could then "hands on" explore how variations in settings and policies will likely impact their behaviors and the quality of their products.

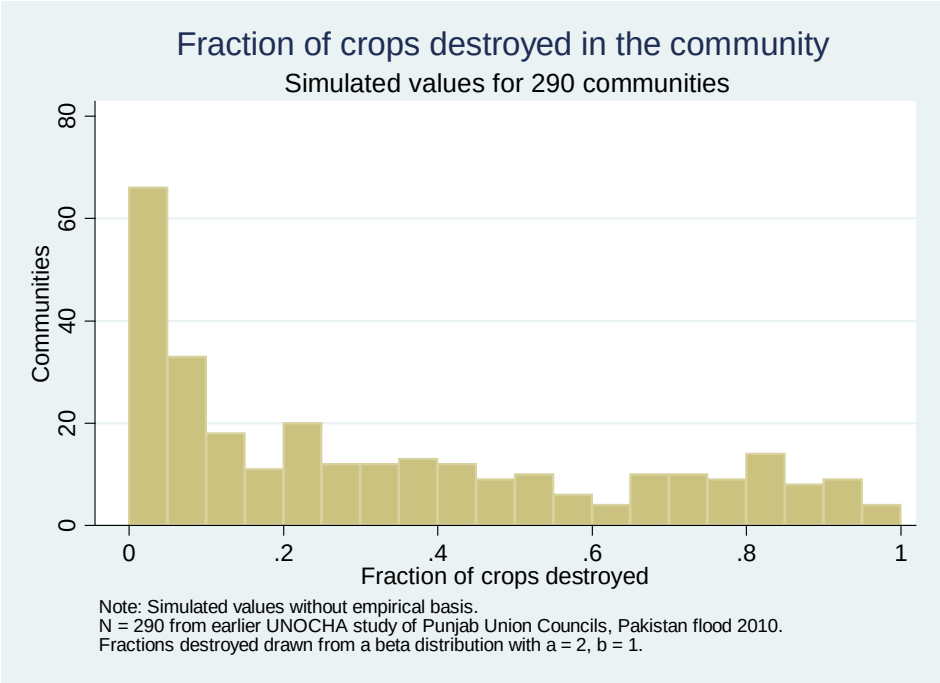
Appendix: Technical considerations

Simulating fractional indicators

Our impact indicators at the community level are fractions, in other words variables defined on the interval [0, 1], or percentages between 0 and 100. Habitually, this type of random variable is modeled as beta distributions (Van Hauwermeiren and Vose 2009: 27; Wikipedia 2011).

Practically, to obtain correlated beta variates, we first generated normally distributed ones. This was done in the statistical software STATA, which offers a routine for this purpose[6]. A STATA logfile, allowing readers to reproduce the same set of random variates, is attached.

Figure 6: Example of the distribution of a simulated indicator



These normal variates are then transformed to uniform ones because the domain of the beta function is the interval (0, 1). We use the same shape parameter values for the generation of variates for all four indicators, alpha = 2 and beta = 1. These transformations, from normal to uniform to beta, can be done in Excel (see workbook). The shape settings make for positively skewed distributions, as seen in this histogram of the fraction of crops destroyed.

The use of identical distributions from which the values of the four indicators were drawn may seem to invalidate the model. It would be easy to rescale the distributions to more

[6] Myerson (2005) supplied an Excel add-in with a function that approximates correlated uniformly distributed random variates. His CORRAND function, however, does not take seeds. For reproduction, therefore, entire random number tables have to be documented.

realistic means for each indicator. As noted in a footnote above, this would then require a reweighting function in the comparisons that determine the impact profile. To avoid this complication in the simulation mechanics, identical distributions are maintained. This poses a challenge of didactic plausibility, not of model validity.

Excel macro

The simulation is directed by an Excel macro programmed in the VBA language that is a standard element of this spreadsheet software. The macro has two essential elements:

- A triple loop structure. The outermost loop regulates runs of assessment itineraries, given desired sample size, and selects the starting point from a random table. The middle loop controls measurements with errors while in a given community and transcribes estimates to results sheets from which their means are collected into a summary sheet for further analysis. The innermost loop calculates distances to all not yet visited communities as well as the ratio of a (key informant- or fly-over-based) measure of physical destruction to a power function of the distance. It selects the community with the highest such ratio to be the next to visit.

- The calculation of distances. Distances are computed between community centroids (which were computed in a GIS application). For projected coordinates (eastings, northings), a simple Euclidean formula works. For the event of working with unprojected ones (latitude and longitude), a Great Circle algorithm with an auxiliary sheet (called "Transfer") has been included as reserve code.

This structure can be adapted for similar simulation purposes. Execution of a run of 50 itineraries took between 5 and 7 minutes, depending on sample size and measurement error inclusion, on a Windows 7 machine with a 40 gigahertz processor and with 8,000 MB memory. This posed a limit to the number of models simulated.

STATA logfile

The key motivation for using STATA is to keep the random variate generation reproducible, by setting the seed in function corr2data below.

```
* Generation of correlated uniformly distributed random variates for simulating disaster
impact variables on Punjab community data set.

* Correlation matrix:
* Copy from Excel, name and define path, e.g.:
* use "C:\...\Correlation matrix from Excel.dta", clear
list
mkmat var*, matrix(C)
matrix list C
clear

* Generate normally distributed variates
* n = 290 is the number of affected Union Councils in Punjab data set.
corr2data x1 x2 x3 x4, n(290) corr(C) seed(1002)
corr x*


gen RecNo = _n
move  RecNo  x1

* Save, e.g. as
* save "C:\..\CorrelatedNormalVariates.dta", replace
```

Supplementary material

The simulation, save the above-described generation of correlated normal variates in STATA, was implemented in an Excel 2007 workbook, with a macro creating simulated itineraries. Using the Excel built-in Analysis ToolPak-VBA function "ATPVBAEN.XLAM!Random" keeps the simulations reproducible.

The workbook may be requested from the author (Benini_Phase2_NeedsAssess_Simul_08Oct12.xlsm). If readers wish to replicate the simulations, the Analysis ToolPak-VBA has to be active (go to Excel Options - Add-Ins - Manage Excel Add-ins: Go, and make sure Analysis ToolPak-VBA is checked. Also macro security has to be set in the menu Developer - Code - Macro Security, to allow all macros.).

References

Benini, A. (2011). Data Management and Data Analysis - Key Issues in MIRA Phase 1 and 2. Washington DC and Geneva, ACAPS.

Fotheringham, A. S., C. Brunsdon, et al. (2000). Quantitative geography : perspectives on spatial data analysis. London ; Thousand Oaks, Calif., Sage Publications.

Garfield, R., C. Blake, et al. (2011). Common Needs Assessments and humanitarian action [HPN Network Paper No. 69, January 2011]. London, Overseas Development Institute (ODI).

Groves, R. M. (2004). Survey errors and survey costs. Hoboken, N.J., Wiley.

Julmy, S. (2011). Coordinating the earthquake response: lessons from Leogane, western Haiti. Humanitarian Exchange. Humanitarian Practice Network (HPN). London, Humanitarian Policy Group, Overseas Development Institute**:** No. 49 [January 2011]. 26-28.

Meindl, J. R., C. Stubbart, et al. (1996). Cognition within and between organizations. Thousand Oaks, Calif., Sage Publications.

Myerson, R. B. (2005). Probability models for economic decisions. Belmont, CA, Thomson/Brooke/Cole.

NATF (2010). Operational Guidance for Coordinated Assessments in Humanitarian Crises [Draft April 10, 2010]. Geneva, UNOCHA and IASC Needs Assessment Task Force.

NATF (2011). Operational Guidance for Coordinated Assessments in Humanitarian Crises [Provisional version, 1st February 2011]. Geneva, UNOCHA and IASC Needs Assessment Task Force.

Pereira, S. C. A. (2009). "Ignorance in a multi-agent setting." Economics Letters **105**(3): 264-266.

Thompson, S. K. and G. A. F. Seber (1996). Adaptive Sampling. New York, Wiley-Interscience.

UNOCHA (2009). Assessment and Classification of Emergencies (ACE) Project. Mapping of Key Emergency Needs Assessment and Analysis Initiatives [Final

Report February 2009]. Geneva, UN Office for the Coordination of Humanitarian Affairs.

UNOCHA (2011). PUNJAB - Humanitarian Assessment Priorities Ranking - Union Councils Level (Total Score) as of 03 January 2011. Islamabad, United Nations Office for the Coordination of Humanitarian Affairs.

Van Hauwermeiren, M. and D. Vose (2009). A Compendium of Distributions [ebook]. Ghent, Belgium, Vose Software.

Wikipedia. (2011)."Beta distribution." Retrieved 29 April 2011, from http://en.wikipedia.org/wiki/Beta_distribution.