Aldo Benini
Patrice Chataigner

A note for ACAPS

Composite measures of local disaster impact - *Lessons from Typhoon Yolanda, Philippines*

27 May 2014

Contents

Tables and Figures

Acknowledgement

Charles Conley and Bekim Kajtazi, both with iMMAP, Washington DC, helped design the maps in this note. Philippe Van Kerm, CEPS/INSTEAD, Luxemburg, answered questions regarding the Betti-Verma algorithm and authorized the reproduction of a section of a technical note in the statistical appendix. Andrej Verity, John Marinos and Joseph Addawe of UNOCHA, Kelly Ryan of UNHCR as well as Sean Ng and Lani Mateo of World Vision reviewed earlier drafts and clarified the antecedents of the prioritization matrices used in the Typhoon Haiyan response.

Summary

Purpose

When disaster strikes, determining affected areas and populations with the greatest unmet needs is a key objective of rapid assessments. This note is concerned with the logic and scope for improvement in a particular tool, the so-called "prioritization matrix", that has increasingly been employed in such assessments. We compare, and expand on, some variants that sprang up in the same environment of a large natural disaster. The fixed context lets us attribute differences to the creativity of the users grappling with the intrinsic nature of the tool, rather than to fleeting local circumstances. Our recommendations may thus be translated more easily to future assessments elsewhere.

The typhoon that struck the central Philippines in November 2013 - known as "Typhoon Yolanda" and also as "Typhoon Haiyan" - triggered a significant national and international relief response. Its information managers imported the practice, tried and tested in other disasters, of ranking affected communities by the degree of impact and need. Several lists, known as *prioritization matrices*, of ranked municipalities were produced in the first weeks of the response. Four of them, by different individuals and organizations, were shared with us. The largest in coverage ranked 497 municipalities.

The matrices are based on indicators, which they aggregate into an index that determines the ranks. Thus they come under the rubric of composite measures. They are managed in spreadsheets. We review the four for their particular emphases, the mechanics of combining indicators, and the statistical distributions of the final impact scores. Two major questions concern the use of rankings (as opposed to other transformations) and the condensation of all indicators in one combined index. We propose alternative formulations, in part borrowing from recent advances in social indicator research. We make recommendations on how to improve the process in future rapid assessments.

Prioritization in Yolanda

For the targeting of the relief, international organizations sent a good number of information specialists to the country, which had already built a comparatively dense pre-crisis statistical environment. ACAPS supported the assessment effort with an extensive Secondary Data Review. While the Review identified gaps in information management, it acknowledged a wealth of accessible data atypical of developing nations.

In the five months following the disaster, response agencies conducted over two hundred needs assessments. The basic prioritization matrix technique was known at least among a section of them. Conceptually, the format was attractive. It offered the double benefit of order and of certainty to find a relevant ranking of the affected populations. At the same time, it was open for additions and revisions. The four matrices that we reviewed share essential features; as much as we can tell, two of them diffused from a common source,

the Global Focus Model (UNOCHA and Maplecroft 2011) while at least one of them was invented independently.

All four combined measures of magnitude, intensity and pre-existing conditions. Magnitude measures are counts of affected units, such as the number of affected persons. Intensity measures are either proportions (such as the percentage of destroyed dwellings) or physical parameters of the storm (such as wind speed). All four matrices expressed pre-existing conditions through the poverty rate. All normalized indicators and sub-indices by means of ranking. Yet these common features did not preclude considerable variation in traits and complexity, coverage and formats, as well as in speed and iterations.

Although the exact primogeniture is impossible to reconstruct, we believe that the four matrices appeared in this order. Their names were not clearly fixed; they varied between file names, named of the key worksheet, or names of the column for the final index. We chose the one that seemed to express the intent of the authors best.

- **Anonymous**: Haiyan Severity Estimates (on or shortly after November 23)
- **Protection Cluster**: Yolanda Priority Vulnerable Municipalities (around December 4)
- **UNOCHA**: Priority Focus Model (around December 11, updated in January)
- **World Vision**: Typhoon Haiyan Affected Municipalities in the Province of Leyte (December 22 [an earlier version had come out on December 1st]).

The main part of this note elaborates on the key characteristics and an appreciation of each of the four matrices. More importantly, a number of common issues have surfaced. Whether they were congenital to the Global Focus Model precursor, or conditioned by the local environment, or both, is not ours to determine. We emphasize those that we believe should be resolved for the benefit of future assessments.

Key issues

What to measure

It remains unclear what the creators of the matrices intended to measure behind the many indicators that they assembled. What was the underlying key concept? We did not find formulated rationales. We see the need to define what it is that we measure, that which ultimately serves as the yardstick for priorities.

Inspired by concepts of vulnerability research, we propose that at the outset of prioritization projects the assessment coordinators design a double model. The *process model* defines, in one simple equation, what we seek to measure as a function of 2 - 4 basic components and of their connections. The *measurement model* indicates how each of the basic components is to be measured. It lists candidate indicators that can be collected with reasonable effort, speed and quality. It suggests a perspective, if not yet the details, for combining them.

To make our notion of process model intuitive, an example of a simple one is:

$$Needs = k * Magnitude * Intensity * f(Pre\text{-}existing\ conditions)$$

where k is an unknown constant expressing proportionality, and $f(.)$ is a function of unknown shape and parameters, and * stands for the multiplication operator. Many others can serve equally well as long as they are simple and appropriate to the type of emergency at hand.

Measurement models must meet three requirements:

1. They must "add apples and oranges" in a satisfactory way.
2. They must produce a final measure that meaningfully relates to the key concept of the process model (e.g., "needs").
3. They must permit hypothesis testing.

A hypothesis of obvious interest is that disasters create relatively few high-impact areas, more middle-impact ones and many low-impact ones. Does the final measure allow us to confirm or refute this hypothetical distribution for the case of Typhoon Yolanda?

Examples of measurement models are worked throughout the main body of this note.

How much to measure

Rapid assessments navigate a magic triangle of speed, cost and quality. This trilemma can force difficult choices between "quick-and-dirty" and information-rich matrix designs. In theory, one can start with a quick-and-dirty product and gradually elaborate. In practice, the arrival of sector-specific assessments may soon render further elaborations of global priorities useless. For example, one of the organizations had built into its prioritization matrix provisions for livelihood indicators, but could not collect them in time for the ranking of municipalities. When these data are ready, they will be useful for livelihoods support, but we doubt that they will be used to revise the general prioritization.

As a thought experiment, we re-calculated two versions of the needs index for the matrix with the smallest coverage - 43 municipalities, all of them in the same province. The quick-and-dirty version used only one indicator each to measure magnitude, intensity and pre-existing conditions. The richer version used a total of eleven. The correlation coefficient between their final measures is 0.72. Data can be expensive; the additional information must have come at a cost to the organization. The information-rich variant calls for greater statistical expertise, which in practice may cause other complications.

Although the results of the information-rich version were more in tune with the default hypothesis, on balance the quick-and-dirty seems preferable. We must not generalize this, but "quick-and-dirty" is a serious consideration.

How to preserve information

Most of the indicators collected in the typhoon response were ratio-level measures. That is, they were variables, such as the number of affected persons, on which all four basic arithmetic operations are legitimate. Reducing ratio-level measures to ordinal ones (ranks) is rarely justified and not at all in the matrices that we reviewed. Ranking destroys valuable information. It blurs the view of the real differences between the hardest hit and the more fortunate.

To make things worse, adding ranked variables creates results that have no valid interpretation. Ranked indices frustrate both decomposition and aggregation. Thus, questions such as *"Are needs in Province X larger than in Province Y?"* cannot be answered with rank-based indices.

One of the strongest messages that this review drives home is: Ranking of ratio-level indicators is detrimental. Ranking destroys information. The resulting final index masks differences that the assessment consumers should be able to see, but are not allowed to.

The right thing to do is to normalize indicators by means that preserve the information, preferably by dividing each one by its sum. The table exemplifies this for just two indicators and five municipalities.

Table 1: Example of normalizing indicators by their sums

Municipality	Affected persons	Affected - normalized	Proportion buildings totally destroyed	Proportion destroyed - normalized
1	54,563	0.11	100%	0.33
2	74,785	0.15	68%	0.22
3	262,856	0.53	27%	0.09
4	62,690	0.13	60%	0.19
5	39,617	0.08	52%	0.17
Total	**494,511**	**1.00**	**307%**	**1.00**

How to measure magnitude, intensity and pre-existing conditions

The four matrices differed in the way they grouped indicators. Two followed an explicit ordering scheme, one on lines anticipating sectoral planning tasks, the other by the presumed utility of each indicator for general response planning. The first arrangement made it difficult to distinguish between magnitude and intensity of local impacts. The second assigned low weights to its "tertiary indicators". As a result, these kept a symbolic presence, but practically were irrelevant. With one minor exception, the matrices aggregated all indicators additively.

A more productive approach will create a sub-index for each major component of the process model - in our example one for magnitude, one for intensity, one for pre-existing conditions. Aggregation to the sub-indices is additive, with weights set by policy or driven by the data. For the latter variant, we demonstrate the use of a redundancy-minimizing algorithm, the Betti-Verma double-weighting rule.

When the sub-indices are aggregated to the overall index (such as of unmet needs), the operation may be additive or multiplicative. Multiplicative aggregation, recently adopted also by the Human Development Index, in many situations is advantageous. It obviates the need for weighting. We present a comparative example in the main text. The multiplicative conforms better to the hypothesis of a skewed impact distribution. Ultimately, the aggregation form should follow the logic of the process model.

We found that in creating intensity indicators the matrix authors tended to divide count variables (e.g., the number of destroyed buildings) by the estimated *affected* population. This unnecessarily discriminates against the more heavily impacted communities. Division by the *total pre-crisis* population is appropriate (division by the affected population discriminates against highly affected communities).

One measure or several?

Every matrix created one composite measure amalgamating magnitude, intensity and pre-existing conditions. This made good sense; the sole or major purpose of the matrix was to rank communities. But is it helpful for assessment consumers who play a role in the detailed response planning and execution? Do planners have an interest in working with separate measures of those components?

For one thing, the components may differ greatly in the degree of uncertainty. In the typhoon, the intensity, measured by such indicators as the proportion of destroyed buildings, predictably decreased with distance for the storm's path. The magnitude was less certain. The estimates that the government put forward of persons affected were perturbed by fluctuating IDP counts and were subject to multiple revisions.

Moreover, there will always be a need to access estimates of affected persons, a key statistic in disasters, rather than simply letting them melt away into some overall index. We believe that overall indices are valuable, but at the same time response planners should be able to see statistics and maps of magnitude and intensity side by side.

Weighting indicators

The matrices differentiated the weights of the indicators that they included in ranking formulas. One organizations justified unequal weights with the strategic emphases that it had already decided for its recovery programs. In other words, policy set the weights. The other matrices lacked stated rationales for the weights chosen.

When weights cannot be derived from known policy, they should be data-driven. Conceptual logic and choice of method matter. When we assume that the construct underlying the index is the *consequence* of the observed phenomena (e.g., "unmet needs" increase with destruction and displacement), we want to cast a wide net. A redundancy-minimizing algorithm such as Betti-Verma is called for (presented in the appendix). By contrast, when the construct is the common *cause* of what we observe in the indicators (e.g., "conflict" destroys property and displaces people), we maximize redundancy. An index may be formed, for example, as the scores of the first factor in a factor analysis. Data-driven weighting methods require some statistical expertise, if only temporarily.

Figure 1 below maps communities by level of unmet needs. The needs scores resulted from a multiplicative aggregation. The sub-indices each were aggregated additively, with weights computed by the Betti-Verma algorithm - magnitude with four indicators, intensity with three, and pre-existing conditions with two (poverty and malnutrition; the latter had not been used in any of the four matrices).

Indicator quality

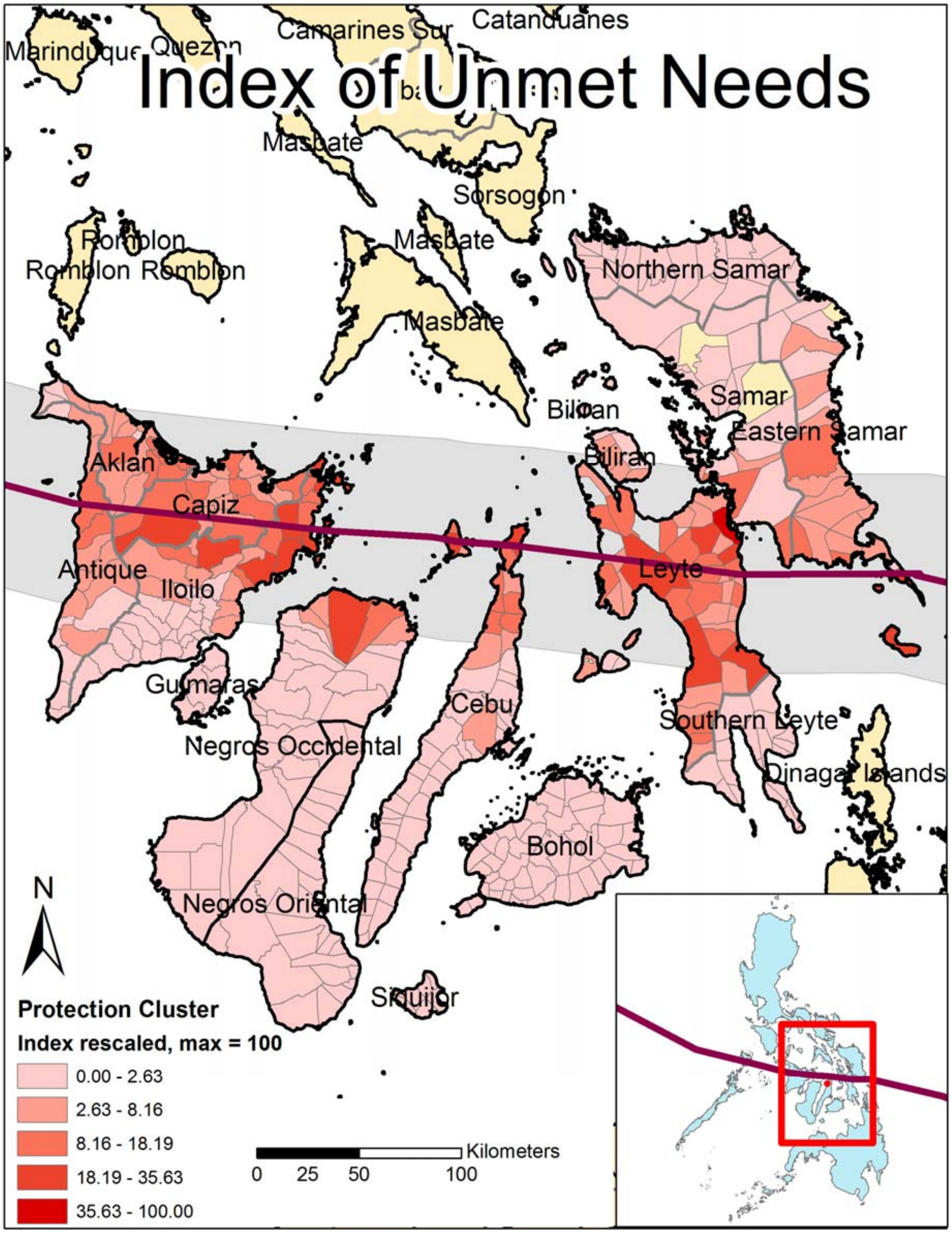
Data quality is a universal concern in needs assessments. Quality has several dimensions, amply discussed in research textbooks. However, in rapid assessments quality competes with cost and speed, two considerations of at least equal importance. The value of an indicator therefore is a function of all three. Decisions to include, collect and use indicators are opportunistic, which also means that they may be revised.

Confidence in the results

Uncertainty is a key concern in humanitarian action and in fact the principal motivation of formal assessments. In the data management itself, however, there is little reflection on the sources and consequences of uncertainty. Measurement error is addressed by updated versions, specification error by including more indicators, and sampling error by full coverage. None of the matrices investigated the robustness of the municipality rankings that they claimed. This is understandable, given time pressure and the difficulty to model error propagation in composite measures.

For a quick-and-dirty model of unmet needs in 408 municipalities, we simulated the robustness both of the ratio-level index scores and of the ranks. For this purpose, we varied the level of measurement error in affected persons and destruction rate. We found that the index had satisfactory robustness. For example, at reasonable error levels, none of the twenty neediest municipalities would drop out of the top quintile.

Figure 1: Typhoon Yolanda-affected communities, by levels of unmet needs



The results are robust not least because of the sharp gradient in intensity between communities closer to, or farther from, the storm's path. If the simulation finding appears trivial, this may not be so in other disasters with more complex impact patterns.

Recommendations

Draw a process model: Define the nature of the final concept that you want to measure, and by which you want to prioritize affected communities. In one equation, write down 2 - 4 basic components and their relationships with the final concept.

Draw a measurement model: List information types that plausibly speak to each basic component. List variables in actual data sets that can be acquired within the constraints of speed, cost and quality.

Acquire the data and evaluate them variable by variable: use histograms and, if feasible, maps; assess outliers and missing values, geographic isolates. Do the same for transformed variables, such as ratios to population.

Choose between "quick-and-dirty" and information-rich designs of the composite measure: Consider "time to market", coverage of each major component by at least one indicator, defensible weights, tolerance among assessment consumers for updated and expanded reports.

If information-rich: For each basic component, build a sub-index. Evaluate relationships between indicators within each sub-index. Determine the suitable aggregation mode. If additive, normalize the indicators by dividing each by its sum. Determine weights, either on policy grounds or with data-driven statistical methods. Evaluate the sub-index through histograms and maps.

Combine the sub-indices in the final index: Determine the aggregation mode, chiefly by what the process model suggests. If additive, normalize the sub-indices so that each sums to one. Choose weights on policy grounds. In the multiplicative model, weights are not needed, but subindices may take an exponent < 1 or > 1 to express smaller or greater importance of a basic component (e.g., stronger poverty orientation).

Edit the final index: For cosmetic reasons, it can be rescaled so that the maximum hits 10, 100 or whatever desired end point. If assessment users expect rankings, at this point ranks are ok, provided the users have access also to the untransformed final index.

Critique, document and share the index and/or its components: Again, use histograms and maps. Evaluate its distribution against the default expectation that disasters cause a small part of the affected area / population to have high values on the dimension that the index measures. A larger part is expected to have mid-range values, the majority to have low values.

Publicize the index in a shape that the assessment users understand, such as in a clean prioritization matrix with appropriate definitions and explanations. Decide whether it is appropriate to present major components separately (e.g., a map of the magnitude side by side with a map of the intensity of the disaster).

Outlook

The prioritization matrices belong in the intersection of decision science and humanitarian action. The practice, in as much as we can observe in the Typhoon Yolanda response, makes an ambivalent impression. Purpose, discipline and creativity are evident, but so is untutored growth, during which no one had the time, inclination or expertise to check detrimental behaviors. Two stand out: lack of definition of what the matrices are intended to measure; the convenience of ranking when in fact the data are strong enough to produce much better than ordinal measures. Both can be corrected.

We believe that the four instances of prioritization tools are but a small part in an evolutionary pool in which ideas and people meet and advance in a process of variation and selective retention. Humanitarian information management has seen rapid progress in areas like telecommunication, mapping and data management. It will be important to periodically rebalance the professional support that ensures both quality and momentum, by increasing access to underused areas of expertise. We hope that this note not only contributes tactical improvements to the practice of this particular tool, but also stimulates more strategic conversations with disciplines like social indicator research, statistics and decision science.

Technical points

The statistical appendix reproduces the detailed Betti-Verma algorithm used to compute weights of additively aggregated indicators. So far it has been implemented in STATA, not yet in MS Excel. We outline an approximation in Excel. Detailed statistical output walks the reader through the process of creating the sub-indices as well as the overall index of needs. For those interested to dig into the robustness simulation, the STATA code is given in full. In the main part of the report, sidebars provide quantitative illustrations of analytic points.

An Excel demonstration workbook is available for download from the same ACAPS Web page. Currently the file name is *Acaps_140527_Philippines_DemoDatatset.xlsx*. The data are a subset of the Protection Cluster prioritization matrix, with sub-index and index variables computed by us. One of the worksheets demonstrates an approximation, easily computed in Excel, of Betti-Verma weights.

Introduction

When disaster strikes, determining affected areas and populations with the greatest unmet needs is a key objective of rapid assessments. The speed, detail and reliability with which this humanitarian intelligence can be assembled depend on a great many factors. Not least among them is the pre-crisis national information environment. These resources are eagerly sought by the information managers that the national and international response brings to the theater. They are then combined with the information collected on the disaster impact as well as on the expanding response. From the combination of pre- and post-disaster sources emerges an operational picture that reflects, over space and time, the magnitude and intensity of the disaster as well as continuing vulnerability, resilience and the beginnings of recovery.

Ranking disaster-affected areas

This note is concerned with the logic and scope for improvement in a particular assessment product, the ranking of affected areas. Typically these are low-level administrative areas, natural settlements or crisis-induced special sites such as camps. The units are ranked by disaster impact or by unmet need, often without an explicit definition of the underlying concept. The function of these rankings is to distinguish areas of greater and lesser need globally. They do not presume sectoral priorities although some of the indicators, by the accident of what information is both relevant and rapidly accessible, are conceptually associated with particular institutional domains.

Several lists of ranked communities were produced in the first weeks of responding to Typhoon Yolanda in the Philippines. We review these for their particular emphases, the mechanics of combining indicators, and the statistical distributions of their final impact scores. Two major questions concern the use of rankings (as opposed to other transformations) and the condensation of all indicators in one combined index. We make recommendations on how to improve the process.

Typhoon Yolanda and information management

Typhoon Yolanda, also known as Typhoon Haiyan, struck the central Philippines on November 8, 2013. By January 7, 2014, the official death toll had risen to 6,183. The disaster affected 14.1 million people, displacing 4.1 million from their homes. More than 1.1 million homes were damaged (Acaps 2014: 1).

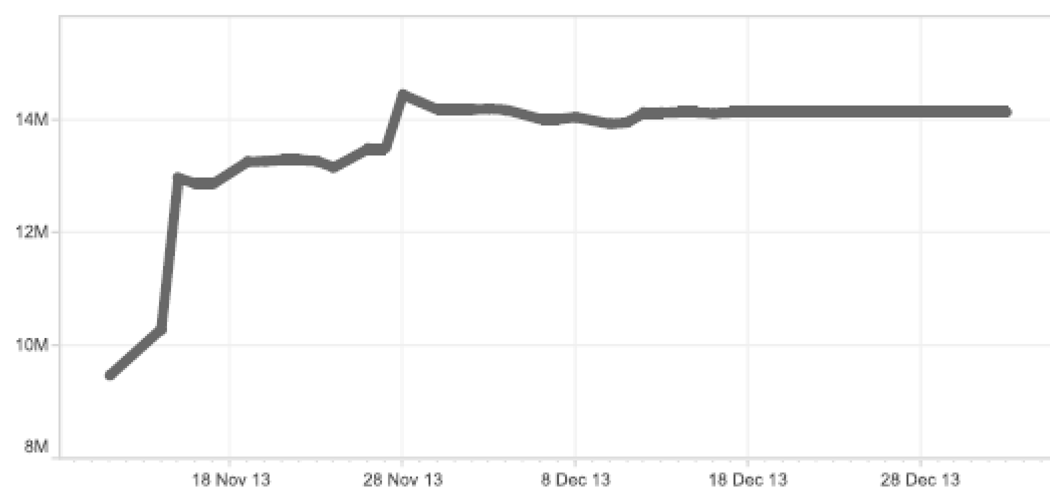
The disaster triggered a significant national and international relief response. The United Nations put out a funding appeal for US$ 301 million, of which some 24 percent were underwritten within a week (Verity 2014: 5). For the targeting of the relief, international organizations sent a good number of information specialists to the country, which had already built a comparatively dense pre-crisis statistical environment (IMWG 2014). These factors accounted for a particular balance of challenges and achievements in needs assessments in this disaster, a situation that may not recur in other places and types of crises, and which limits the scope of our findings and recommendations.

Secondary Data Review

ACAPS contributed to the effort through an extended Secondary Data Review (SDR), the final version of which it published in January (op.cit.). The wealth of information defies brief review. Nevertheless, two specific points warrant mention. First, the SDR provides a detailed list of information gaps and needs, together with recommendations for future assessments. The tone of these is for more information and more data. This is in conflict with modern decision theories that stress bounded rationality, the cost of information and the need for succinct heuristics under time pressure. This note, while looking also at potential additional data types, favors community rankings based on less and faster information.

Second, the SDR documents, in a novel and painstakingly calculated timeline, the evolution of estimates of affected persons that the coordinating government authority publicized between November 11 and the end of 2013. Figure 16 on page 13 (op.cit.) shows how the estimated totals moved in steep leaps and on almost flat plateaus over a two-week period, and then gradually settled to a final figure. Another graph depicts an even sharper rise in the number of affected villages and neighborhoods (known as *barangays*), followed by a significant downward correction from around 13,000 to 11,000. These fluctuations, together with the wide margin of interpretation of "affected persons", caution against the inclusion in community rankings of "affected persons" on an equal footing with better defined and more precise impact measures.

Figure 2: Trend of affected population figures (Figure 16 in the Acaps SDR)



The number of affected persons has been an element of the community rankings that are the main topic of this note. The rankings were produced as part of what the needs assessment community calls "prioritization matrices" (NATF 2014: for background and a repository of documents). In this regards, one of the SDR recommendations reads (p. 3):

> An overall Prioritisation Matrix should be discussed and validated among actors (including agreed definitions of key terminology such as “affected”, “partially” or “completely”, etc.), and should be disseminated and updated regularly. The Prioritisation Matrix could take the form of an overall dashboard and be made available at regional and provincial level (currently, 26 provinces have reported affected populations). The dashboard should be shared every month and should incorporate all available data provided by the different governmental bodies involved in the response. The matrix should then be supplemented by available secondary data. Two Prioritisation Matrixes are currently available, one designed by the Protection Cluster and the other by UNOCHA.

ACAPS also had access to samples of such matrices from two sources outside UN agencies. We will review these as well.

Dynamics of humanitarian response

The prioritization matrices are products of the humanitarian response, which evolves rapidly in the days and weeks after the disaster onset. The matrices themselves change, not only because more data becomes available, but also because the types of decisions that direct the response change. Proxy indicators of impact such as wind speed and storm surge in this typhoon may be remotely estimated almost immediately. They give way to impacts observed within the social fabric. After the initial magnitude and intensity have been mapped, sector-specific indicators grow more important, as can be seen in two of the four reviewed matrices. It is beyond the scope of this note to describe this co-evolution in detail. Instead we narrowly focus on the logic that connects indicators with the final ranking of affected communities, and on the critique of the ranking method itself.

[Sidebar:] What data to expect in sparser statistical environments?

The response community in the Philippines benefitted from the relatively dense pre-existing statistical environment. In future prioritization exercises in other countries, data availability may be more problematic. It is realistic to assume that some, but not all of these indicators can be produced with reasonable coverage.

Table 2: Types of data likely available in affected-community prioritization

Pre-existing	In crisis
Population, area, population density	Affected population
Low lying/ flood prone areas	Internally displaced persons
Past food insecurity rates	Food insecure persons
Malnutrition rates	Totally destroyed houses
Poverty rates (prevalence, less likely depth and severity)	Partially destroyed houses
Previous shocks (from some kind of crisis)	Partially destroyed schools/classrooms
Number of disabled persons	Totally destroyed schools/classrooms
Previous conflict/security incidents	School dropouts
	Humanitarian access
	Responder organizations working (3W)
	Is an area with no data (as an uncertainty measure)

Besides the coverage, which is readily assessed, the reliability will be vary from situation to situation, and from indicator to indicator, and will be harder to assess.

Prioritization matrices

A Google search on "prioritization matrix" returns nearly 40,000 hits. Adding "humanitarian" nets 714 documents, suggesting a recent, yet by now firm reception in this institutional realm. In the wider world, the matrices come in numerous formats and procedures. The matrix format that NATF is promoting is more consistent. Essentially it is a platform for multiple attribute decision making (Yoon and Hwang 1995). The spreadsheet format presents affected units in rows, and indicators of their situation or needs in columns. The indicators, some in transformed (categorical or ranked) shape, are weighted and additively aggregated in an index (also called a composite measure or total score). The index may again be ranked, ultimately producing a league table of affected units ordered along an impact or unmet needs dimension. The approach generally is deterministic; fluctuations are incorporated via updates and/or limited experiments with different weights. Measurement error, omitted variables and alternative functional forms are, to our knowledge, not concepts taught in the guidance documents or factored in in field practice. Of course, the users of these tools do take into account the flux and uncertainty of the post-disaster evolution. The key advantages of the matrix are order and the certainty that it will produce a relevant ranking of units while always open for additions and revisions.

Order of appearance

In the wake of Typhoon Yolanda, four prioritization matrices came to the notice of ACAPS (two of which are mentioned in the SDR). Although the chronology is difficult to reconstruct (the versions saved by ACAPS may not be the first publicized), we believe that they emerged in this sequence:

- **Anonymous**: Haiyan Severity Estimates (on or shortly after November 23)
- **Protection Cluster**: Yolanda Priority Vulnerable Municipalities (around December 4)
- **UNOCHA**: Priority Focus Model (around December 11, updated in January)
- **World Vision**: Typhoon Haiyan Affected Municipalities in the Province of Leyte (December 22 [an earlier version had come out on December 1st]).

The matrices differ not only by publication dates, but also by coverage (number of municipalities), format and, to some degree, intent. Their intellectual precursors are known only in part. UNOCHA's and those of the Protection Cluster can be traced back to the Global Focus Model, a world-wide country-level vulnerability ranking (UNOCHA and Maplecroft 2011: and earlier). Some of its devices were adopted in the Philippines in ranking exercises in the response to the Ketsana / Ondoy flood disaster in the Greater Manila region in 2009 (Wikipedia 2014b). After Haiyan, officers at UNICEF and UNHCR in Manila agreed on a common model inspired by the Global Focus Model. By contrast, the World Vision team in Leyte developed their model independently. The antecedents of the Haiyan Severity Model are not known.

Anonymous: Spatial model

Attributed to a Philippino student in Spain, Miguel Antonio Garcia, this early matrix exploited public data immediately available after the disaster, chiefly the distance from the storm path and the height of the surge (the latter from a map in the New York Times!). UNOCHA subsequently published a map based on these results.

Key features

The matrix has records of 175 affected municipalities. The ranks are based on severity scores with four arguments: pre-disaster poverty incidence, population density, distance from the storm path (in four categories), and height of the storm surge. The formulas of the spreadsheet have been replaced with their values, but one can guess that the raw variables were transformed to the interval [0 1], using different formulas. Poverty incidence was "adjusted" for population, and there is an interaction term between distance to the storm path and surge height. The severity score is the mean of the weighted inputs.

Appreciation

This initiative is noteworthy for its early date and for the use of spatial and socio-economic data then available from public sources. The coverage is smaller than in the subsequent assessments; it is also heavily biased towards areas close to the storm path. Thus it resembles more of a purposive sample.

Since the formulas are not shown, the operations leading to the score are not entirely transparent. The author's notes suggest that the score is meant to reflect magnitude (via population density), pre-existing conditions (via poverty incidence) as well as intensity (via the storm indicators). The aggregation combines interactive (probably multiplicative) with additive operations. This choice is neither described nor explained.

In sum, the first matrix was early, parsimonious, ingenious in the use of spatial data, but not as transparent as one might wish.

Re-analysis with a different aggregation

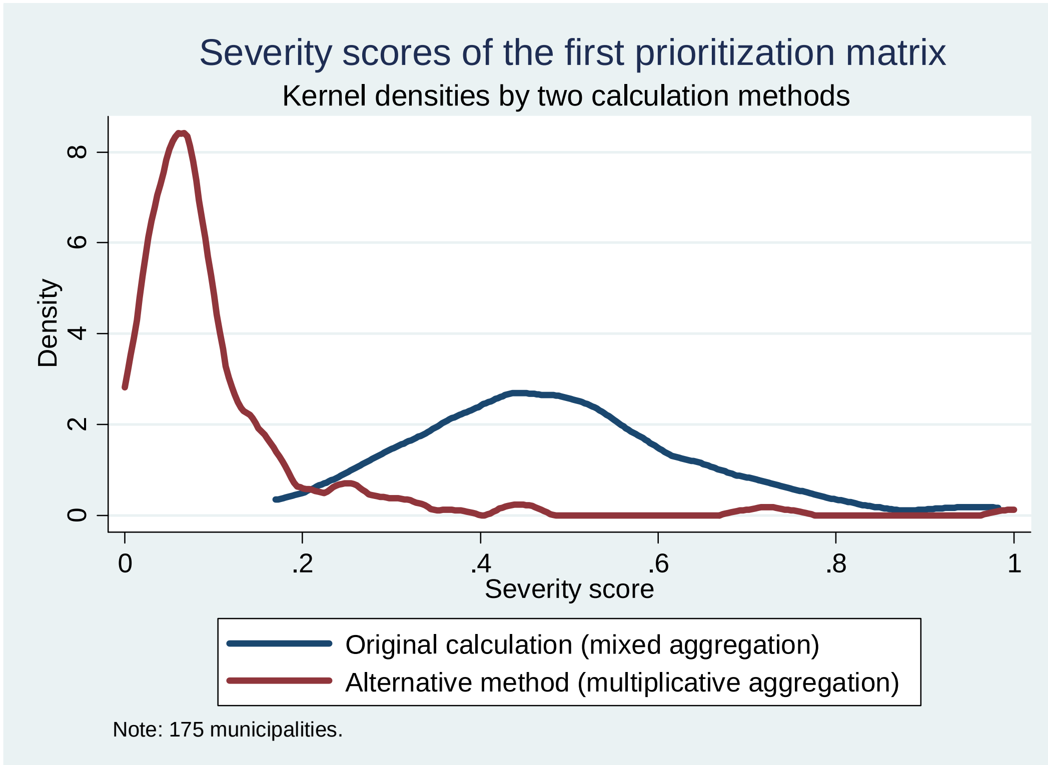
We therefore take the liberty to look at the same data with a different method of combining them. The figure below contrasts the distributions of the impact scores by two calculation methods. The first is by the anonymous author, using mixed aggregation operations. The second is by us and is largely multiplicative:

*Score = Population density * poverty incidence * (0.75 * distance category + 0.25 * surge height)*

(the last term with weights is as in the original) and subsequently normalized by the minmax method. In other word, the original stresses additive aggregation, which makes it a "compensatory method". More storm surge can be compensated, for example, by lesser population density. In qualitative terms, the logical operator for this is the "OR".

The alternative scoring is dominated by multiplicative aggregation. Compensation is less readily available. To reach the same score when, for example, the surge height is halved, some other factor will have to be doubled. In qualitative terms, the logical operator is the "AND".

Figure 3: Spatial model - distributions of severity scores by aggregation method



The results are vastly different. The additive method, by its inbuilt compensation, tends to concentrate scores in the middle. With the multiplicative method, positive values on all indicators and high values on at least one are necessary conditions for high scores. It concentrates scores in the low range. This concurs with the intuition that disasters tend to create relatively few high-impact areas, more middle-impact ones and many low-impact ones. But it can create the impression, if inappropriately presented, that the overall impact was mild.

Ranking or real differences?

This dilemma will accompany us through the rest of the note. It can be mitigated by the simple ranking of units, replacing the scores. In fact, in terms of order, the two methods agree largely. Spearman's rank order coefficient between the scores of the two methods is a high 0.78. The question then becomes whether one should sacrifice the interval-level information[1] and settle for mere ranks. Ranking blurs the view of the real differences between the hardest hit and the more fortunate. Interval and ratio-level scores are harder to explain, heightening the need for explanation at a time when nobody has much time.

[1] If the reader wonders why interval level: The minmax normalization to the interval [0 1] reduces the measurement level from ratio to interval.

Protection Cluster: Vulnerable Areas

Key features

The Protection Cluster produced a "Total Score" on a variable that is not conceptually named in the spreadsheet, but which may easily be understood as the local disaster impact. The Cluster scored 408 affected municipalities. The area scored is the one shown in the map in the summary.

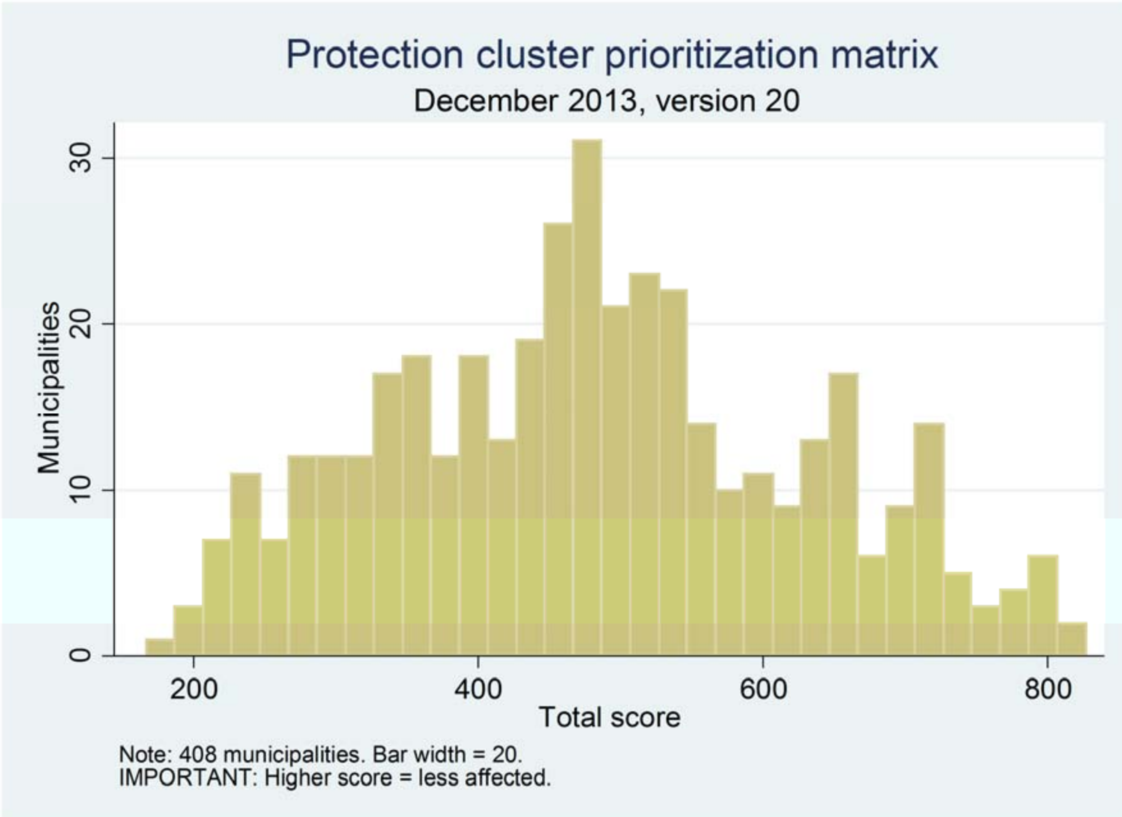
The score is a composite of primary, secondary, and tertiary indicators, with each group combined in a sub-index. A sub-index is the sum of the *ranks* that a municipality attained on the concerned indicators[2]. The three sub-indices have decreasing weights (0.6, 0.3 and 0.1) and are aggregated additively. Based on the total score, the municipalities are assigned an "overall rank".

- The **three primary indicators** comprise the ratios of IDPs and of damaged houses to the affected population as well as the poverty incidence (the latter presumably from a survey taken before 2013).
- The **five secondary** indicators comprise the ratio of the affected persons to the estimated 2013 population as well as the number of villages / neighborhoods (*barangay*) affected by armed opposition groups or inhabited by indigenous populations[3].
- The **three tertiary** indicators refer to government programs - whether the government ranked a municipality among the 171 top affected, as well as the number of villages / neighborhoods in which two conflict and peace programs (PAMANA and NAPC) were active.

Because of the ranking of indicators, the total-score distribution congregates in the center. The distribution nevertheless is not quite normally distributed - not only because it is theoretically bounded (ranks cannot be smaller than 1 or larger than the sample size), but also because of the several spikes in the upper range.

[2] To be precise: The field rank, in which the highest value receives the first rank.
[3] Known as the CADT program, for "Certificates of Ancestral Domain", besides other programs such as the "Certificates of Ancestral Land Title" (CALT) (http://www.iccaregistry.org/en/countries/4).

Figure 4: Distribution of the total scores in the Protection Cluster matrix



Appreciation

The matrix produced by the Protection Cluster went through more than twenty iterations. Our data are from version 20. Its original contribution is the consideration of special disadvantages that municipalities exposed to armed conflict and those with indigenous populations have been suffering from times before the disaster. The scoring formula also acknowledges the presence of certain government programs aimed at bringing peace and development to particular areas.

With eleven indicators (and the estimated 2013 population as a twelfth, used as a denominator of the affected population), the resulting index is complex. The Cluster transformed the individual indicators to rankings, presumably to make them more comparable and to give them equal influence before weighting. The use of these rankings raises many questions; these are difficult to answer in the absence of some explicit rationale by the authors.

With the exception of binary variables designating the simple presence or absence of certain government programs, the indicators are all ratio-level count variables. Admittedly, some have few distinct values (the number of New People's Army-related incidents, for example, takes only 14 values between 0 and 17). But still they are count variables with high information specificity. Do they warrant ranking?

Is there a valid case for ranking?

Generally, ratio-level variables are transformed to rankings when the only reasonable assumption about the relationship with an underlying variable of interest (such as unmet needs) is that it is monotonous[4]. Surely, such an assumption is too modest for some of the indicators that the Cluster employed for the score, notably destroyed housing and affected populations. Assume that we have three municipalities A, B and C. In A, 10 percent of the population are affected, in B 20 percent and in C 80 percent. In terms of ranks, C is 1st, B 2nd, A 3rd. Few would argue, though, that the difference between A and B should have the same consequence for the score as that between B and C. However, rankings exactly do that - they create equal differences between adjacent values.

This is a first important point that ought to be considered in future impact scoring situations. Is ranking really the best way of normalizing indicators? Ranking invariably destroys some of the information. Other normalizing functions such as dividing the indicator by the sum of its values preserve it (while also giving each indicator the same importance before weighting). The objective in preserving the original differences is to verify the plausible distribution of few high-impact, more medium-impact and many low-impact units - or its surprising deviation therefrom.

Creative elements

The second question is related to the creative part of the Protection Cluster scoring - the inclusion of populations-at-risk and of government programs. The Cluster allowed most of these to contribute to the total score with low weights only. The rationale for this is far from obvious.

The first at-risk group is the poor, captured in the poverty rate. The way this rate is connected to the scoring function expresses a policy preference. An *additive* form in the aggregation implies that poverty is an independent additional consideration. Depending on the weights, a highly affected municipality with a low poverty rate would receive a similar score as a mildly affected one with a high poverty rate, all other things being equal. In the ultimate consequence, the resources provided for Yolanda victims would serve both disaster rehabilitation and targeted economic development, with only an accidental connection between the two.

[4] Conceptually, this would take us to the neighborhood of fuzzy-set approaches, where standardized rankings are occasionally used to express the membership function. See, e.g., Longest and Vaisey (2008). In Excel, the function *percentrank* does this kind of transformation. But there are no strong reasons to fuzzysize these indicators. If untrustworthy outliers are the problem, there are other methods to control them that do not waste as much information as ranking does.

The *multiplicative* form has different consequences. The contribution that the ratio of affected persons to the population makes to the score is multiplied by the poverty rate. In the limit case, when there are no poor, the contribution is zero. In policy terms, the non-poor would have to take care of themselves. On the other hand, municipalities with a high ratio of affected persons and a high poverty rate would be on top of the relief priorities (again, other things being equal!). This expresses a belief that the unmet needs of the disaster-stricken poor are particularly acute, and added up in municipalities with many poor, particularly large.

Technical standards for data analysis cannot decide such policy questions. They can only highlight them.

Technical points

A thorny point is the technically correct handling of the conflict and special government program indicators. Here, due to the small numbers of distinct values, rankings are particularly uneven. In the extreme, the NAPC program indicator has only two values, 1 for being present in a municipality, 0 for its absence. The rank function assigns the value 1 to participating municipalities, and 314 to those outside the program. Instead of working with ranks, dichotomizing all these variables (with a reasonable threshold for NPA-related incidents) and then simply adding them into a sub-index seems much preferable. The ratio of affected persons should be kept separate.

Finally, a remark is due on the denominator that was used in a number of normalized indicators. The numbers of IDPs and of persons associated with destroyed and damaged homes were divided by the number of *affected* persons. This discriminates unnecessarily against municipalities with high proportions of affected persons. Dividing by the *total* municipal population seems more appropriate.

Mere symbolism?

In sum, the Protection Cluster formula is innovative up to a point. It does bring into the equation groups at special risk, and it endeavors to give added influence to the municipalities where they live. However, these gains are minimal, due to the particular distributions of the conflict and special-program variables and to the low weights given to the secondary and tertiary sub-indices. A deeper statistical analysis reveals that the conflict and special program indicators account only for about one percent of the total score variance. At best, they are symbolic.

[Sidebar:] What influence do the indicators have?

We are interested to see which indicators most strongly determine the so-called total score. Because they were ranked, the score no longer is a simple linear combination of the raw indicators and, in a statistical regression model, is not fully explained by them. So, how much do the particular indicators explain?

To simplify this question, we compressed seven conflict and program indicators into two principal components[5]:

- These **conflict and special government** program components, plus
- the ratios of **IDPs** and of
- persons with **destroyed and damaged homes** to the affected population,
- the **poverty** rate,
- the ratio of **affected persons** to the population, and
- being among the **171 top municipalities**,

jointly account for 93 percent of the total score variance[6].

A fair apportioning of the explained variance to the indicators can be obtained employing a technique known as "Shapley value decomposition" (Royston and Kolenikov 2013; Shorrocks 2013). In this table, we regrouped the indicators slightly, by their associations with disaster impact vs. with special protection considerations.

[5] Technically, by using polychoric principal components (Kolenikov and Angeles 2004). We retained the first two components, which between them account for 66 percent of the variability in the seven indicators.
[6] The dataset shows only 160 municipalities as part of the 171 top priority ones.

Table 3: Total score of the Protection Cluster matrix - decomposed

Indicator	Variance explained if this is the only explanatory variable	Shapley value decomposition of variance in the joint model
DIRECT DISASTER IMPACT:		
Ratio **IDPs** to affected population	50.76%	26.67%
Ratio persons with **destroyed or damaged homes** to affected population	54.45%	22.79%
Ratio **affected** persons to total population	3.86%	4.93%
Government-defined **top 171** municipalities	34.51%	11.65%
SPECIAL PROTECTION CONSIDERATIONS:		
Ratio **poor** persons to 2013 population	35.91%	26.03%
Conflict and special programs, component 1	0.38%	0.66%
Conflict and special programs, component 2	0.66%	0.36%
Residual	-87.43%	
Total variance explained, joint model	**93.10%**	**93.10%**

The results are telling. Individually, IDPs and physical destruction have the greatest influence on the score. Because they are strongly correlated, in the joint model their variance shares are roughly halved. Together, they still account for nearly half of the variance. The poverty rate takes up another quarter. The ratio of affected persons to the total population contributes minorly. The government top 171 municipalities' effect on the total score individually is strong; Shapley value-decomposed, it accounts for much less, for approx. a tenth of the explained variance. The conflict and special-program indicators are negligible in the greater scheme of variance shares.

In terms of prioritization matrix design, the lesson is clear: either transform and weight the indicators such that they contribute significantly, or omit them. Avoid symbolism.

UNOCHA's Priority Focus Model

Key features

UNOCHA computed a "Priority Focus Model" score or "PFM" for 497 affected municipalities. The PFM is a weighted sum of *ranked* indicators. It incorporates information on affected persons, damaged houses (partially vs. totally), the poverty rate and, as a denominator, the municipal population in 2010. Affected persons and damaged houses are used twice, in their amounts and in their ratios to population. The poverty rate is included additively. The PFM thus takes seven arguments. The weights are 1.25 for the two affected person-related ranks, 1.75 for the totally damaged house-related ranks, 0.75 for the partially damaged ones, and 1 for the ranked poverty rate.

The ranks were calculated using the Excel function *percentrank*, whose values fall into the interval [0, 1]. The theoretical range of the PFM is [0, 8.5] (2 * 1.25 + 2 * 1.75 + 2 * 0.75 +1 = 8.5); the observed values range from 0.18 to 7.40.

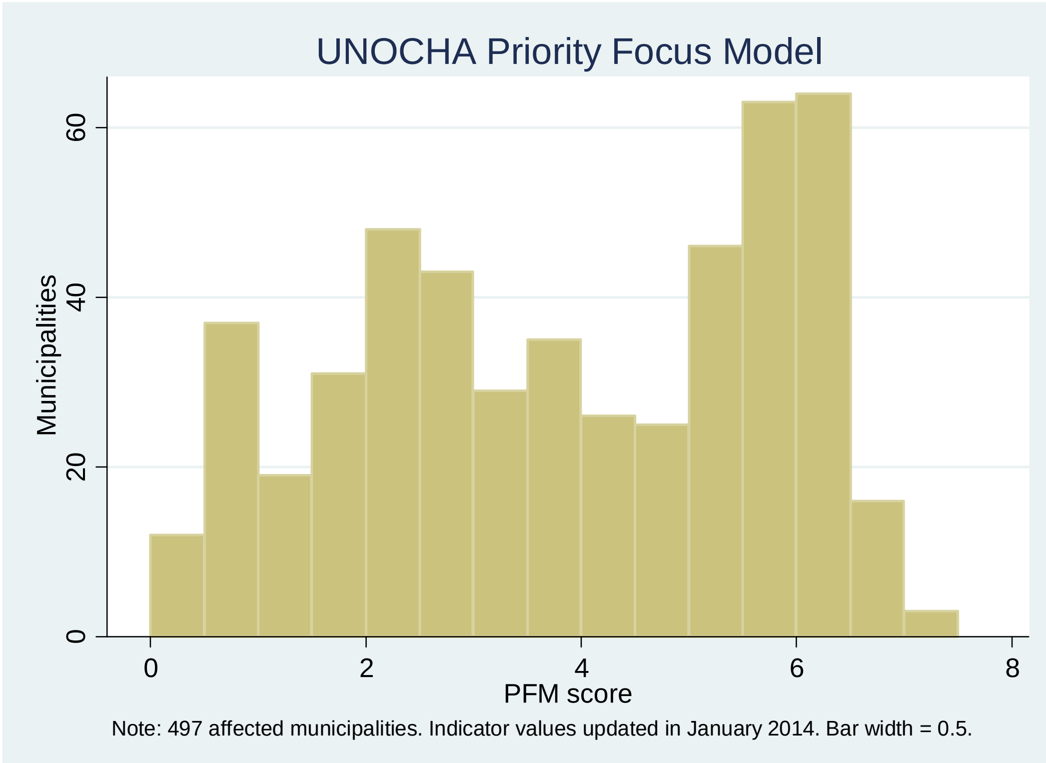
Appreciation

The PFM strikes a balance between magnitude and intensity, by combining the (ranked) amounts as well as ratios to population of its impact indicators. It biases the score to poorer communities by the additive poverty rate term.

The distribution of the score is clearly bimodal[7]. With two peaks, it has somewhat better discrimination. Clearly there is a massed group of high-impact municipalities, and another group that was less affected. If we set a value in the trough, say 5, as the cutpoint, almost 40 percent (192 out of 497) municipalities are in the upper group. Although this is not substantively justified, a 40/60 partition is a useful first cut as the prioritization summary. There must have been processes at work creating this two-peak distribution of impact, although we do not know what they were.

[7] Some, looking at the histogram, may say: even trimodal. However, a three-component normal distribution finite-mixture model (output not shown here) does not clearly assign the members of the third component to the latent classes.

Figure 5: UNOCHA Priority Focus Model, histogram of the score



Questions and speculations

Nevertheless, there are serious questions about the underlying model, of which, as far as we know, the assumptions have not been stated explicitly. The reader therefore may forgive us if some of what follows is speculative:

- **How were the weights determined?** The only obvious rationale is that the weights for partially damaged houses should be smaller than those for completely damaged ones. One consideration *may* have been to weight by the certainty of impact: Destroyed homes were precisely counted and indicated high impact (weight: → 1.75); partial damage is an elastic notion and also indicates less severe impact (→ 0.75). "Affected" persons may have been estimated broadly, but in response to all kinds of impacts, thus justifying some mid-sized weight (→ 1.25).

- **What about the weight of the poverty rate?** Depending on municipal sample sizes, the sampling variance in these rates may be considerable. Yet, since this is an indicator conceptually distinct from the immediate disaster impact, the unit weight chosen for the PFM is as good or bad as any other weight. It might be easier to defend the inclusion of the PFM if at first an intermediate index were formed of the impact indicators only, and then, in a second step, *multiplying* this index by the poverty rate. This would sidestep the problem of weighting the poverty rate. It would intuitively reflect a commitment to heighten attention to the poor among the disaster-affected communities.

- A more radical question - more radical because it breaks the mold of one composite measure - is this: **Should magnitude and intensity be expressed in <u>one</u> measure?** Or should they rather be kept separate? Perhaps with a map for each side by side, and with summary tables displaying them in adjacent columns? We will elaborate a possible rationale for this separation in the final section.

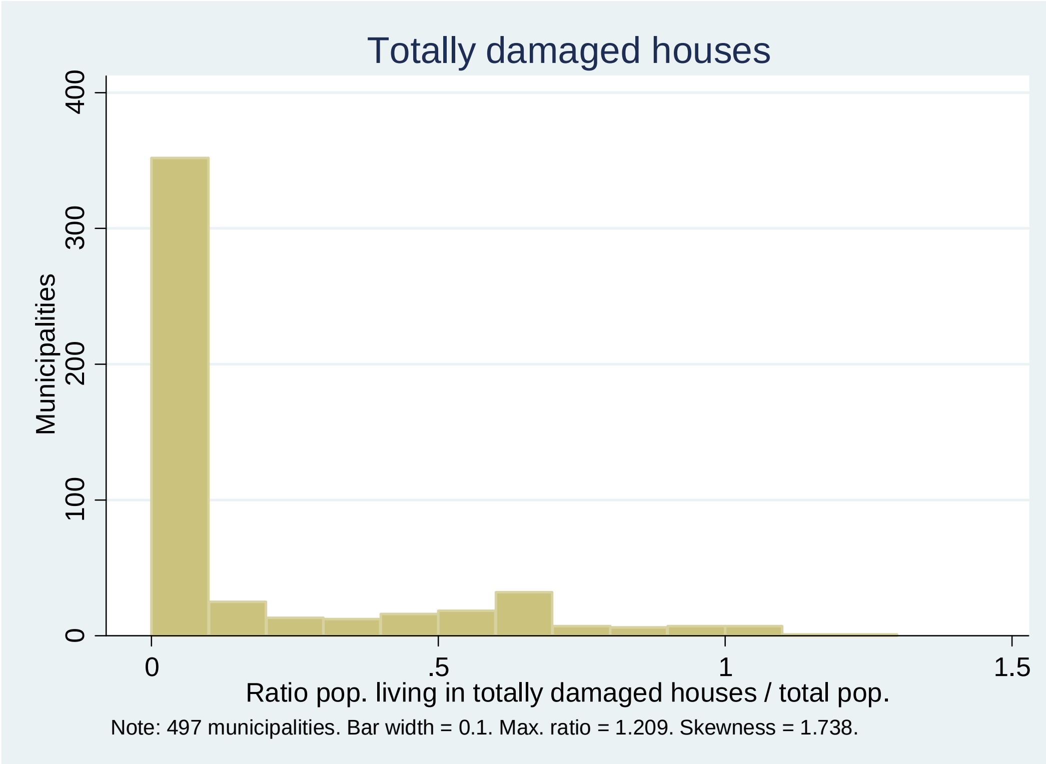
- **Is the type of normalization chosen appropriate?** The PFM combines *ranked* indicators. In other words, it adds (weighted) ordinal variables. This raises two serious objections:
    - **Information loss:** The drawbacks of ranking were mentioned before (page 21); ranking sacrifices information and blurs differences. The loss of information can be considerable, as the sidebar below calculates for one of the indicators. Types of normalization that preserve information should be preferred. Dividing by the sum of indicator values does so. Since, by design, the sum of any thus transformed indicator is one, they all have the same importance before weighting. If decimal fractions do not look attractive, the normalized indicator values can all be multiplied by some magnitude of 10, as cosmetically desired, without any substantive distortions.
    - **Quantities without valid interpretation:** Ordinal variables cannot be added meaningfully. Therefore indices formed as weighted sums of ranks have no valid meaning. This is the strongest argument against the use of ranks.

  Both objections are pursued in the sidebars below.

[Sidebar:] Information loss due to ranking

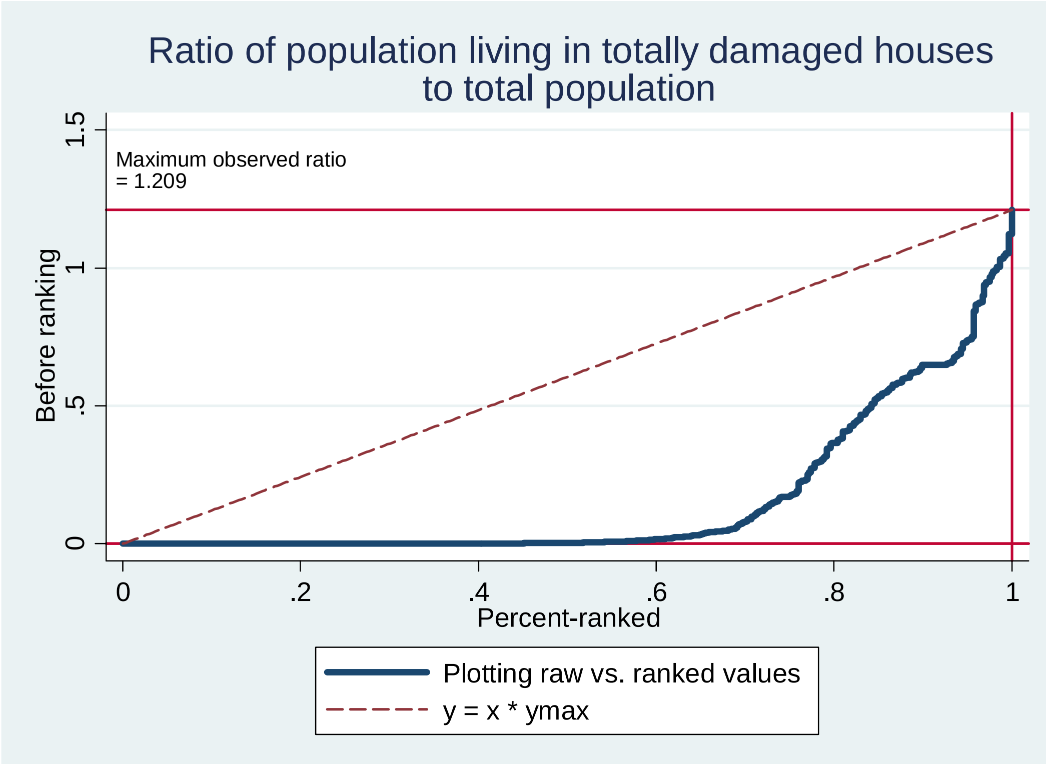
Some of the indicators have highly skewed distributions. The skewness is not a drawback. In fact, it is valuable information because it tells us something about the quantitative relationships between highly and less impacted units. For demonstration, we select, among the indicators included in the PFM, the ratio of persons associated with totally damaged houses to the total population. In nine municipalities, the ratio exceeds 1 although not by very much. We leave these as are.

Figure 6: Ratio persons in totally damaged houses to total population - histogram



The skewness is lost in the ranking. Ranks, by definition, are uniformly distributed over their range (i.e., in the case of Excel's *percentrank* from 0 to 1). The graph below plots the original ratio against its ranked values. While the original values (through their order) predict their ranks completely, the ranks predict the original values poorly. One way to visualize the information loss is to think of the distance between the dashed line and the plotline as an overestimate. The information loss could then be expressed as the area between the dashed and the plotline divided by the area of the triangle. This value is about 0.74, and by this (unorthodox) measure the loss thus about 74 percent.

Figure 7: Damaged house ratio - plotting raw vs. ranked



The demonstration is mainly didactic[8]. A more conservative loss estimate is based on the product-moment correlation (*correl* in Excel). In analogy to the R2 in regression models, and the unexplained variance part as 1 - R2, we measure the loss due to ranking as 1 - corr^2. We find that corr(raw ranked) = 0.7233; hence 1 - corr^2 = 0.477. By this measure, nearly half of the information is lost.

By either measure, the loss due to ranking is considerable. Ranking thus needs solid reasons - plausible benefits that offset the information loss. Except for the convenience of a ready Excel function (*percentrank*), these benefits are not obvious in the PFM case.

[Sidebar:] Why ranked variables cannot be added meaningfully

Why indicators should not be ranked and then added (weighted or not) is best demonstrated with a small demonstration from a totally different field - lotteries.

[8] Statistically, it would be more correct to standardize raw values and ranks to their z-scores (i.e., subtract the mean and divide by the standard deviation), then plot the difference of their z-scores against the mean, producing a so-called Limits-of-Agreement graph (Lin 1989). This is less straightforward, however.

Three friends, A, B, C, participate in three consecutive lotteries. Their gains are shown in the upper panel of this table. These financial gains are ratio variables.

In the lower panel, the gains are ranked for each lottery. An index is then formed by adding the ranks over the three lotteries. The friends receive their overall ranks on the basis of this index.

Table 4: Didactic example of ratio-level vs. ranked indicator aggregation

Lottery gains					
Friends	**First lottery**	**Second**	**Third**	**Total gain**	**Overall rank**
A	$ 1,000	$ -	$ -	$ 1,000	1
B	$ 100	$ 10	$ 10	$ 120	2
C	$ 40	$ 20	$ 20	$ 80	3
Ranked gains					
Friends	**First lottery**	**Second**	**Third**	**Mean rank**	**Overall rank**
A	1	3	3	2.33	3
B	2	2	2	2.00	2
C	3	1	1	1.67	1

It is obvious that the overall ranking is completely reversed, depending on whether it is formed on the total gains, or on the averaged ranks in each lottery. The overall rank based on the already ranked gains is misleading. It has no meaningful interpretation.

The same holds for rankings based on already ranked indicators in prioritization matrices when the original variables are ratio-level.

World Vision's "Overview of Affected Municipalities"

Key features

World Vision scored and ranked 43 municipalities, all of them in Leyte Province[9]. The scoring workbook presents a detailed rationale in its main sheet and score guide sheet. The architecture of the workbook is clearly oriented toward response decision-making, after some strategic decisions on emphases had already been taken. *"Shelter will be the main focus [..] As such, it has been weighted higher than the other sectors"* (Score Guide). In the calculation of the "Overall Rank", shelter has a weight of 0.35, affected population, IDPs, and education each have a weight of 0.10. Livelihood has the same weight as shelter, but the needed data were not yet available in this version. The ranking thus was preliminary.

[9] The island of Leyte is shown on the map in the summary.

Interestingly, the final scoring sheet presents not only the "Overall Rank" for the municipalities, but also an "Aggregate Score". While the Overall Rank is the sum of the weighted sectoral ranks, the Aggregate Score results from the *unweighted* sum of the sectoral scores.

Among the sectoral features, three seem particularly noteworthy:

- Like the other matrices, World Vision's combines magnitude and intensity. Differently from the others, it combines them through multiplicative terms. For example, the "affected population severity score" was computed as "Total No. Affected Population * % Population Affected * Poverty Incidence" (sheet "Aff").

- While the poverty incidence is multiplicatively linked, it impacts only the affected population score. In the overall ranking, this component commands a small weight (0.1).

- The shelter score results from a combined damage and gap analysis, based on damage figures as well as estimates of the emergency and recovery shelter gaps.

The workbook has numerous other sheets. Some report practical actions already taken; these are summarized in a dashboard sheet ("Sum"). For each municipality, a data collection form displays the types of information needed under the recovery/rehabilitation plan.

Appreciation

World Vision's matrix is compelling by its neat presentation, explicit and detailed rationale, as well as the parallel computation of scores and ranks. It is outspoken about data gaps, with provisions to update the scoring and ranking once IDP and livelihood data become available in a more complete degree. At the same time, the authors warn of manpower shortages that limit constant updating. The thrust of the exercise is visibly towards the practical response, reporting also the relief distributions and other activities achieved to date.

When scores are not real scores

There are nevertheless some dilemmas in the general-impact scoring, which is the section of the workbook that concerns us here. "Score" (in the sector sheets), "aggregate score" and "overall score" (in the "Score" sheet) are misnomers. De facto, these score variables are track ranks - rankings that give the highest number to the highest ratio-level input. The "ranks", by contrast, are field ranks, with the highest number given to the lowest input. To make this clear, here is an excerpt of the shelter rank and score, for four municipalities, based on the "shelter and damage gap score" (which is a real score):

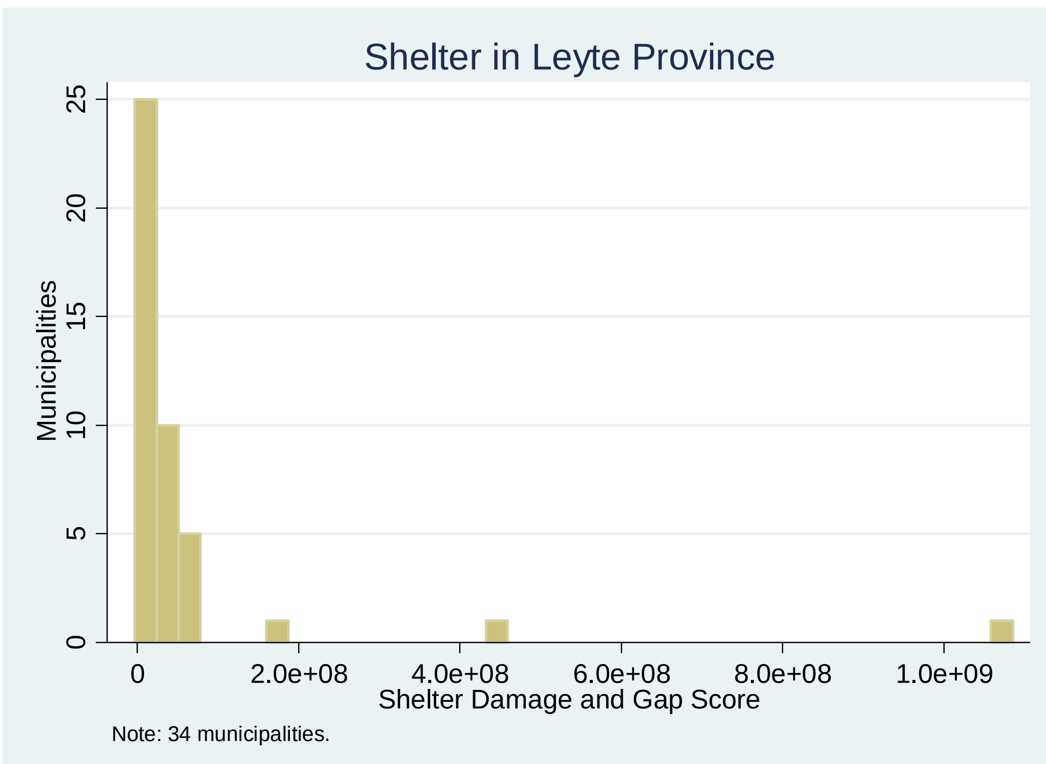
Table 5: "Rank" and "score" in the World Vision matrix

Shelter Damage and Gap Score	Rank	Score
72407304.00	5	9
72169363.75	6	9
39628246.19	12	8
3406163.46	37	1

Note: From sheet "Shel"

The way this matrix is scored destroys information, as ranking of interval- and ratio-level variables always does. The product-moment correlation between the basis (the "Shelter Damage and Gap Score") and the "Score" is a mere 0.46.

Figure 8: Distribution of the shelter damage and gap score, Leyte Province



The default expectation: Few / more / most

The histogram confirms the plausible default distribution: a few highly impacted units, more medium-impacted ones, a majority of low-impacted ones. Three highly impacted ones are distinct outliers. The authors hint at *"anecdotal evidence of over-reporting of losses and damages"*. The constructive response is, not to rank the entire set, but to treat the outliers specially:

- If they are exaggerated, but plausibly still worse affected than the rest, allocate more assessment resources to them.
- If their worse condition is not plausible, set their scores equal to the highest non-outlier[10].

Finally, the poverty rate. The multiplicative term that World Vision used for it is the preferable option. It eschews the problem of separately weighting poverty, as noted before. However, in the mechanics of this matrix, the poverty rate got stuck in a low-weighted subordinate score. The rank order correlation coefficient between the rate and the final Overall Rank is a mere 0.23, statistically insignificant. If the poverty orientation were to be more than symbolic, the multiplication should take place at a higher level, such as by multiplying the aggregate score by it.

In sum, this workbook is a feast for the eyes. It is well presented. It is well explained. It connects with detailed response planning. It nearly achieved, then shied away from, its potential to work out a proper overall score. Such as score would preserve the information in the raw measures, rather than reduce it by ranking. A ranking could still be produced, if found necessary for planning and resource mobilization, but it would happen only at the very end, on the basis of the unranked overall score.

Magnitude, intensity and pre-existing conditions

All four prioritization matrices produce composite measures that combine indicators of magnitude, intensity and pre-existing conditions. The latter are measured with one indicator - the pre-crisis poverty rated. Mechanically, the four proceed in similar ways. The indicators are normalized by ranking; the ranked indicators are weighted; the weighted indicators are aggregated by simple addition. Some matrices re-rank the aggregate score.

The need for a stronger rationale

While the purely mechanical aspects - the spreadsheet formulas - are transparent in most aspects, rationales for them are absent. We have to guess them. Convenience of Excel's ranking functions, weighting in terms of perceived relative importance, spreading the risks of overlooking something important by including as many indicators as available may have motivated the choices.

This creates three problems:

---

[10] A variety of one-sided winsorizing (Wikipedia 2014c).

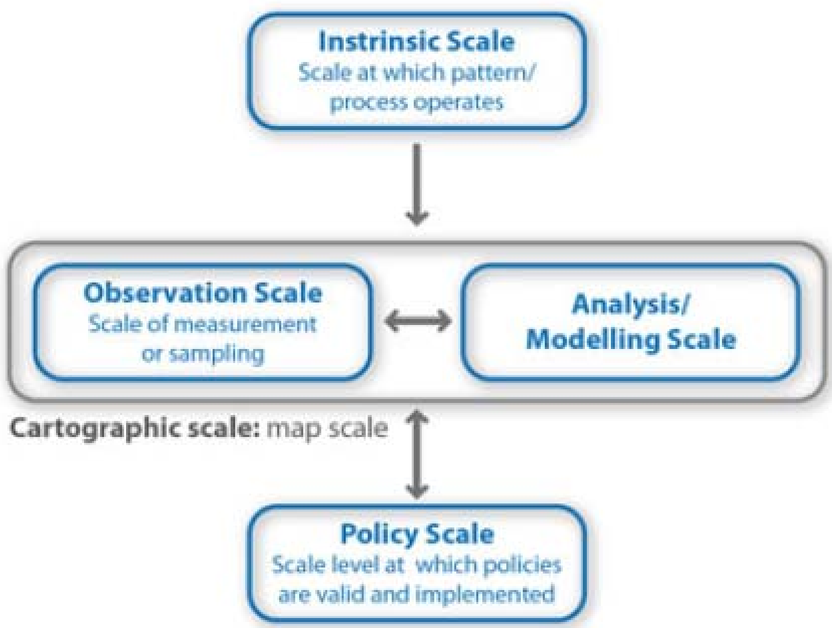
1. Why indicators are included, and how much each contributes remains unclear.
2. Some indicators are strongly correlated, but essentially measure similar things. This introduces a form of double-counting.
3. Magnitude, intensity and poverty all get fused in one expression while the response planners will have to consider them separately.

This section explores approaches that mitigate those problems. We illustrate them with statistics of the World Vision matrix data. Thanks to its small number of observations (43 municipalities), some of the relationships can be depicted in scatterplots. The same logic applies, of course, to assessments with larger samples, but their visualization would be more challenging.

Conceptual assumptions

At first we need a conceptual model. In this context it is helpful to observe how vulnerability researchers have been thinking about populations exposed to risk. In conceptualizing spatio-temporal "vulnerability cubes", Kienberger et al. (2013: 1347) observe how phenomena of interest interact at three levels. What they call the "intrinsic scale" in this diagram means the inherent dynamic of the process in point - here Typhoon Yolanda and its aftermath. Human observation detects and deciphers it, to a degree. Although Kienberger's notion of scale is rooted in human geography and not primarily in needs assessment, this meta-scheme, if you will, helps to organize our own perspective.

Figure 9: A model of models of the assessment process



Process vs. measurement models

We can translate this scheme for the needs assessment process. The important point is to distinguish between two models - the model representing the *intrinsic process* ("Analysis / Modelling Scale") and the model of *measuring* objects pertaining to it ("Observation Scale"). Their relationship is shaped by the expectations that the assessment consumers have in terms of useful output, indicated by the double-pointed arrow connecting the "Policy Scale" box.

The practice of prioritization matrices is driven by efforts to systematize the observation side, with little explicit attention devoted to the intrinsic-process modeling size. This section aims to redress the balance. We begin by proposing a simple model of the unmet needs in the affected population.

A simple model of unmet needs

We state the assumptions about our object, the unmet needs of the people in the disaster zone as follows:

1. The needs are proportionate to the magnitude of the disaster, other things being equal.
2. The needs are proportionate to the intensity of the disaster, other things being equal.
3. The needs increase with the pre-disaster adverse conditions, other things being equal.

This leads to a multiplicative model of the underlying variable:

$$Needs = k * Magnitude * Intensity * f(Pre\text{-}existing\ conditions)$$

where k is an unknown constant expressing proportionality, and $f(.)$ is a function of unknown shape and parameters. The model is multiplicative because in an additive form of any kind

$$Needs = a*Magnitude + b*Intensity + c * f(Pre\text{-}existing\ conditions),$$

a one-percent increase in one of the variables would not entail a one-percent increase in the needs.

f(Pre-existing conditions) reflects the sensitivity of the unmet needs to the pre-existing conditions. Among these, poverty may be the most consequential element. A fair assumption is that disaster compromises poor individuals, households and communities in their basic needs more severely than their better-off neighbors. The degree to which needs assessment formulas factor in available poverty information should, in theory, determine the poverty orientation of the disaster response. Other factors, such as conflict potential too may be important.

Measuring the concepts

The reasons for choosing a multiplicative model of unmet needs in the process model are different from those that argue for or against multiplicative aggregation in the measurement model.

On the measurement side, multiplicative aggregation recently has received increasing support in the social indicators community. The reasons are in part substantive, in part technical. For example, the Human Development Index (HDI), after twenty years of additively aggregating life expectancy, education and GDP sub-indices, switched to multiplicative aggregation in 2010. Here the concern was substantive. The multiplicative scheme was adopted because

> *“'the more severe the deprivation on any dimension, the more difficult it is to have a high HDI. This better addresses UNDP’s concerns about focusing on the state of the more vulnerable segments of society in determining the level of human development in any country'”* (Tofallis 2013: 1329, quoting from a review report).

Others have emphasized technical benefits. The multiplicative model does not need weights or a conversion factor (it *can* take weights, in the shape of powers of the variables). By contrast, the additive model needs either weights or dimensionless normalization from the start. Normalization and weights are needed in order to add apples and oranges.

The multiplicative model simply multiplies apples and oranges, leaving the interpretation of the new synthetic fruit "applange" (or "orapple") to the substantive model side (the right-side box in the above diagram). Another argument in favor of multiplicative aggregation is that the information loss is generally smaller in such indices than in those formed additively (Zhou and Ang 2009).

Precautions when multiplying

In defense of the prioritization matrix, we make two reservations. First, most published indices with multiplicative aggregation make use of the geometric mean or of unequal weights in the shape of exponents. In our case this would mean

$$Needs = k * Magnitude^{a} * Intensity^{b} * Pre\text{-}existing\_conditions^{c},$$

with $a + b + c \leq d$, and with d most often set = 1, such as in the geometric mean with $a = b = c = 1/3$. Certainly, one can discuss, on the *measurement* side, the wisdom of dampening the effect of the measured poverty rate with an exponent $c < 1$. A lower c expresses a weaker poverty orientation. However, there are no valid grounds to let $a < 1$ or $b < 1$ in the *conceptual* view of how disaster magnitude and disaster intensity interact in creating needs. The geometric mean is not appropriate for our purposes.

Second, some indicators speak to the magnitude, others to the intensity of the disaster. Absolute impact figures (mostly) express aspects of magnitude; relative figures, denominated to pre-disaster population or assets, are better at capturing intensity[11].

Accordingly, we combine some indicators to form a magnitude sub-index, and others to form an intensity sub-index. If we have more than one indicator of pre-existing conditions, we form a sub-index for this domain as well.

In the design of these sub-indices, additive aggregation may be preferable. Some participating indicators may read zero in some units. In a multiplicative sub-index model, a zero value even in one indicator will result in zero for the subindex value for the given unit. This is not appropriate. Think of a situation where public buildings are generally built more strongly than the average private residence. The schools in some municipalities may not have suffered significant damage, but private residences have. Thus if "classrooms damaged" is one of the indicators included in the magnitude index, municipalities reporting zero classrooms damaged would be assigned a zero magnitude under the multiplicative model, but will show positive values under the additive model. Clearly, the latter is more appropriate - in this limited function.

Stepping back for a moment

Our discussion so far assumes that it is a good thing to represent unmet needs in just one number (per assessed locality). For the purposes of response planning, this is not compelling. One can work with one composite measure that combines magnitude, intensity as well as (some function of) pre-existing conditions. Alternatively, one can make a case for keeping them separate (or for combining intensity and pre-existing conditions, but keeping magnitude separate). We will revert to this later, but it is important to know that we have options.

Sub-indices: Quick-and-dirty vs. rich information

Regardless of that strategy, there are tactical decisions to be made about the design of the magnitude, intensity and pre-existing conditions sub-indices:

- Should we work with a minimalist set of indicators, in order to minimize normalization, weighting and aggregation issues?
- Or should we make use of the maximum number of (reasonably complete) indicators, employing more advanced methods to take care of redundancy?

To investigate these questions, we proceed as follows. We produce two versions of the sub-indices and resulting needs index:

[11] With obvious exceptions. Wind speed, an absolute figure, expresses intensity in a typhoon-type disaster.

- We implement the quick-and-dirty approach by assuming that magnitude, intensity and pre-exising conditions each can be sufficiently expressed in just one indicator. We take the number of affected persons for the magnitude, the fraction of totally destroyed houses for the intensity and the latest official poverty rate for the pre-existing conditions.
- For contrast, we take the information-rich approach to the sub-indices of magnitude and of intensity. As regards pre-existing conditions, we only use the poverty rate, as in quick-and-dirty[12].

Using the World Vision data, we combine municipality-level figures of

- `Persons affected`
- `Totally damaged housing units`
- `Partially damaged housing units`
- `Total IDPs`
- `Total students affected`

in the magnitude index, and figures of

- `Percent persons affected`
- `Percent totally damaged housing units`
- `Percent partially damaged housing units`
- `Ratio of IDPs to population in 2010`
- `Percent students affected`

in the intensity index. The indices are computed with the help of an algorithm that minimizes redundancy of indicators in composite measures. It is built around the so-called "Betti-Verma double-weighting rule", which is

> *"sensitive to both the relative frequency of items and the correlation among items. The correlation is taken into account so that two perfectly correlated items 'count as one' and only two uncorrelated items fully 'count as two'"* (Pi Alperin and Van Kerm 2009: 2).

The Betti-Verma formula is presented in the statistical appendix. To our knowledge, the algorithm has not been implemented in Excel. We outline a simplified version in Excel.

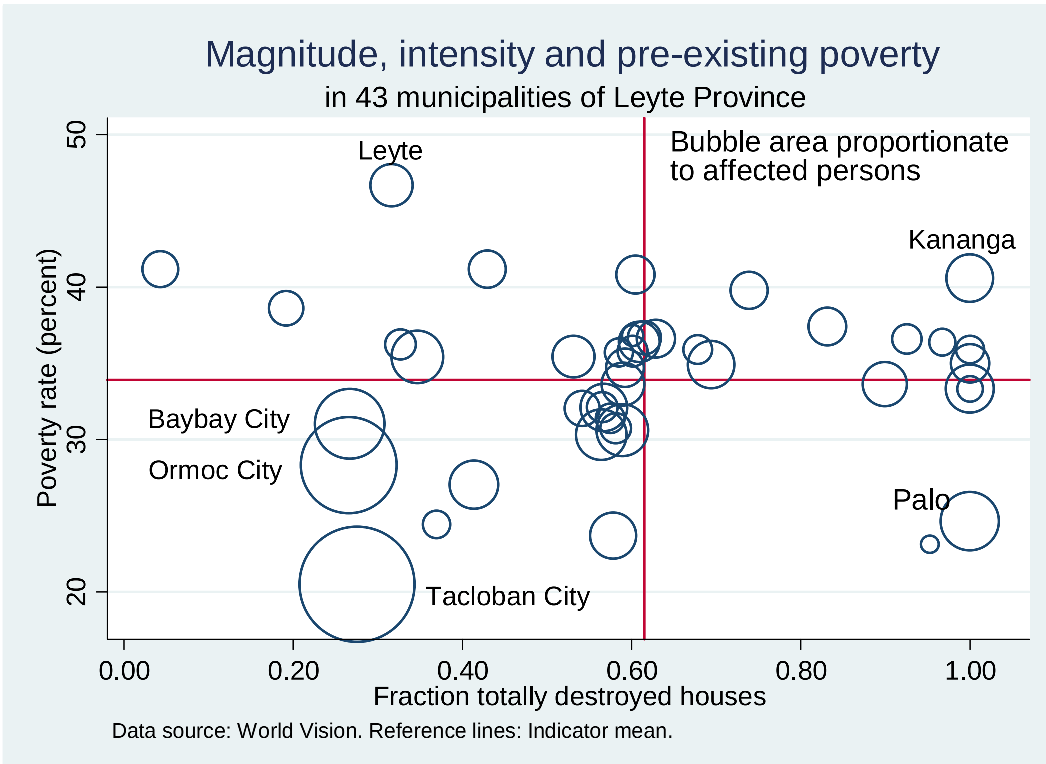
[12] The Philippino school authorities measure the nutritional status of new elementary student cohorts. These data are accessible, but have not been incorporated in any of the four matrices. At the time when we reviewed the World Vision matrix, we did not have these data in readily usable form (we later used it in another experiment - see page 70 in the statistical appendix.

Some readers may expect more background on the poverty rate. We are not competent to offer this for the Philippines. For the complexities of poverty in this country in a social indicators perspective, see Bayudan-Dacuycuy and Lim (2013).

Results of "quick-and-dirty"

This bubble plot keeps the three dimensions separate.

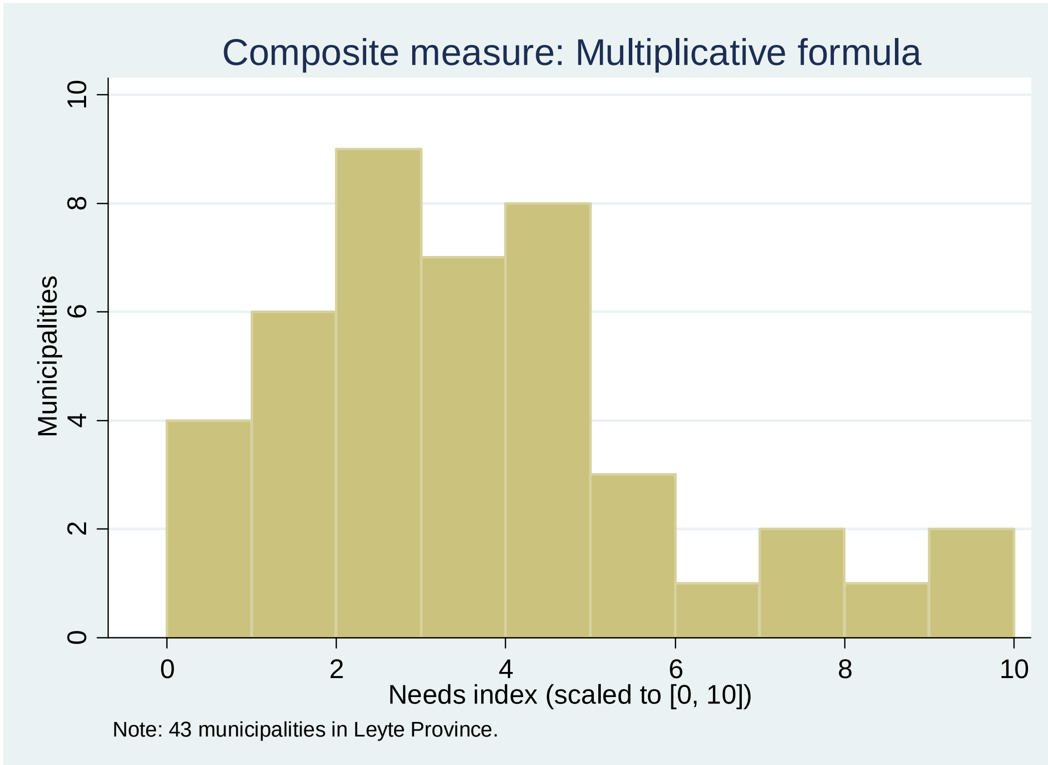
Figure 10: Magnitude, intensity and poverty in Leyte Province



A full interpretation of the chart is not needed here. It is useful to the point that we see right away that the larger cities were neither particularly poor, nor heavily destroyed. These two features are likely causally linked: cities were better off and as such had a higher proportion of sturdy buildings. A large cluster of municipalities is close to the averages of destruction and poverty. A smaller cluster near the Kananga marker is composed of high-destruction, above-average poverty communities with small and medium numbers of affected persons.

At least, there is some structure in this three-dimensional distribution. What preoccupies us naturally is whether a composite measure would reflect anything of it in an interesting way? To probe, we create a simple multiplicative aggregate index: *Affected persons * fraction totally damaged houses * poverty rate*, and watch its distribution.

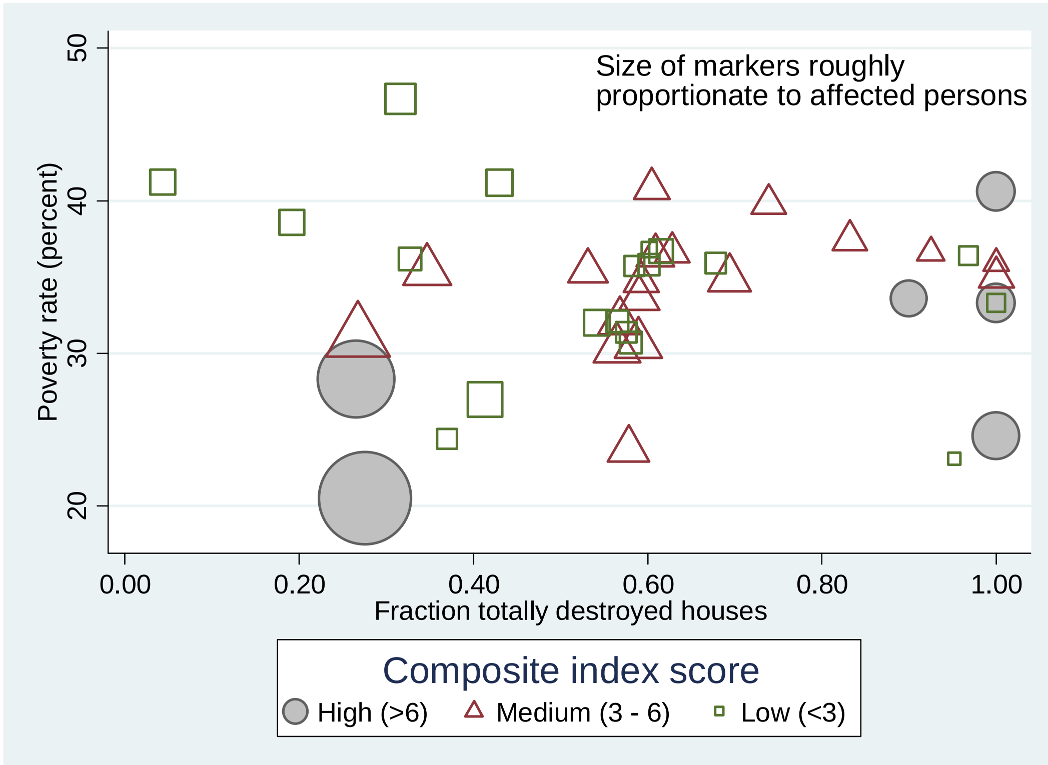
Figure 11: Composite measure, multiplicative formula - histogram



The resulting distribution satisfies our expectation that few units are highly impacted. However, at least in this case, it displays more mid-impact units than low-impact ones, for which the particular exposure of the Leyte island may be responsible.

We classify the 43 municipalities by high, middle, and low needs levels and visualize them in the scatterplot built on the same indicators as the previous.

Figure 12: Needs score in relation to three variables



It is obvious that the index confounds two very different groups of affected municipalities in the high-needs category. Two larger urban communities made it there. Four municipalities with high housing destruction, but fewer affected persons qualify as well[13].

While the composite measure may be valid in terms of the total extent of unmet needs, it obscures important compositional information that the response planners should have. If we find it useful to combine magnitude, intensity and pre-existing conditions in one measure, then the users should have convenient access also to the three components. Measures of magnitude, intensity and pre-existing conditions should be presented in a table and/or in maps side-by-side.

Results in the information-rich approach

As listed above, we use five indicators to build the magnitude sub-index. After we have normalized them, the Stata procedure *mdepriv* (Pi Alperin and Van Kerm 2009: op.cit.) calculates adjusted weights. Indicators with the highest average correlation with all other indicators and/or with a low coefficient of variation receive the lowest weight. "Students

[13] An additive formula, using normalized indicators that sum to 1 and with unit weights, barely changes the result when the high-needs range is widened to [4.5 - 10]. However, this depends on weights whereas the multiplicative formula does not.

affected" behaves more uniquely than the others and therefore winds up with the highest weight. In this way, double-counting of indicators measuring similar things is minimized, and the breadth of the phenomena looked at is incorporated more amply.

Table 6: Weights of the indicators forming the magnitude sub-index

Indicator	Weight
Persons affected	0.107
Houses, totally damaged	0.069
Houses, partially damaged	0.238
Total IDPs	0.212
Students affected	0.375
Total weight	**1.000**

Note: Variables normalized to sum to 1.
Students affected by damaged classrooms.

We do almost the same for the intensity sub-index. However, a precaution is due because the percentages of houses totally damaged and partially damaged have a high negative correlation (-0.82). This is so because the sum of the two cannot exceed 100 percent. Including both these indicators in the procedure yields a high negative weight on the percentage of partially damaged houses. Although the weights still sum to one, they cannot be interpreted[14].

We thus exclude the partially damaged houses and work with only four indicators.

Table 7: Weights of the indicators forming the intensity sub-index

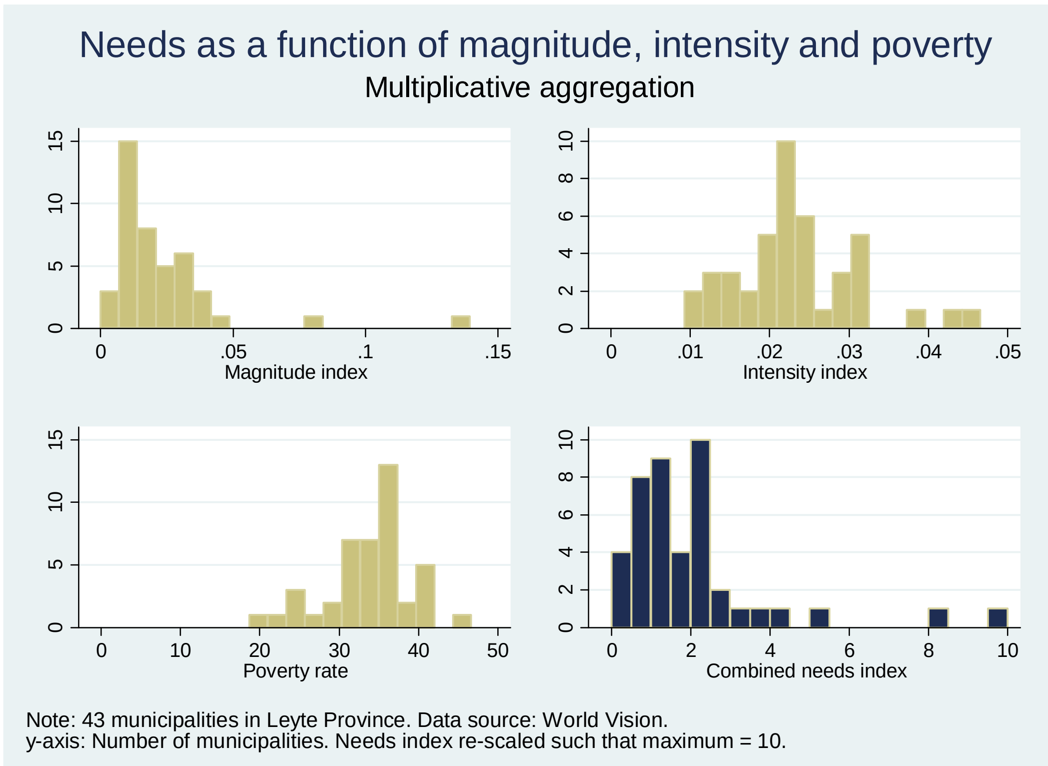
Indicator	Weight
Percent houses totally destroyed	0.187
Percent persons affected	0.039
Ratio IDPs to population	0.424
Percent students affected	0.350
Total	**1.000**

mdepriv penalizes "Percent persons affected" severely, because of its low discriminating power. Most of the 43 municipalities were declared having 100 percent affected persons. The coefficient of variation is 8 - 10 times smaller than those of the other three indicators. "Percent houses totally destroyed" is penalized chiefly because of its higher average correlation with the others. The procedure ensures the broadest non-redundancy.

[14] This problem is not due to the specifics of housing damage, but to the nature of the variables as compositional data. It will occur with any compositional variable of which more than one category are used in the index formation (such as "percent totally damaged" as well as "percent partially damaged"). For background (and software to remove the effects of auto-correlation), see Thió-Henestrosa et al. (2005).

Now we are ready to combine the information-rich versions of the sub-indices with the poverty rate. Again we proceed by multiplicative aggregation.

Table 8: Combined needs index



Under this model, the populations of two municipalities - Tacloban City (score = 10) and Kananga (8.4) - have aggregate needs far outstepping all others. There are four municipalities in the score range 3 - 6, and 37 municipalities in the range below. Tacloban City got to the highest score on account of its high magnitude, Kananga because of higher intensity and elevated poverty. Tacloban City's magnitude score was high despite the low weight given to affected persons. The city had high values on partially destroyed houses and on IDPs, both of which are relatively highly weighted.

What follows from this finding?

Three questions immediately arise from this result:

1. In terms of a "policy scale", is the needs index **useful**?
2. Assuming it is useful, is it **robust** to measurement error?
3. Is it significantly different from the index of the **quick-and-dirty** method?

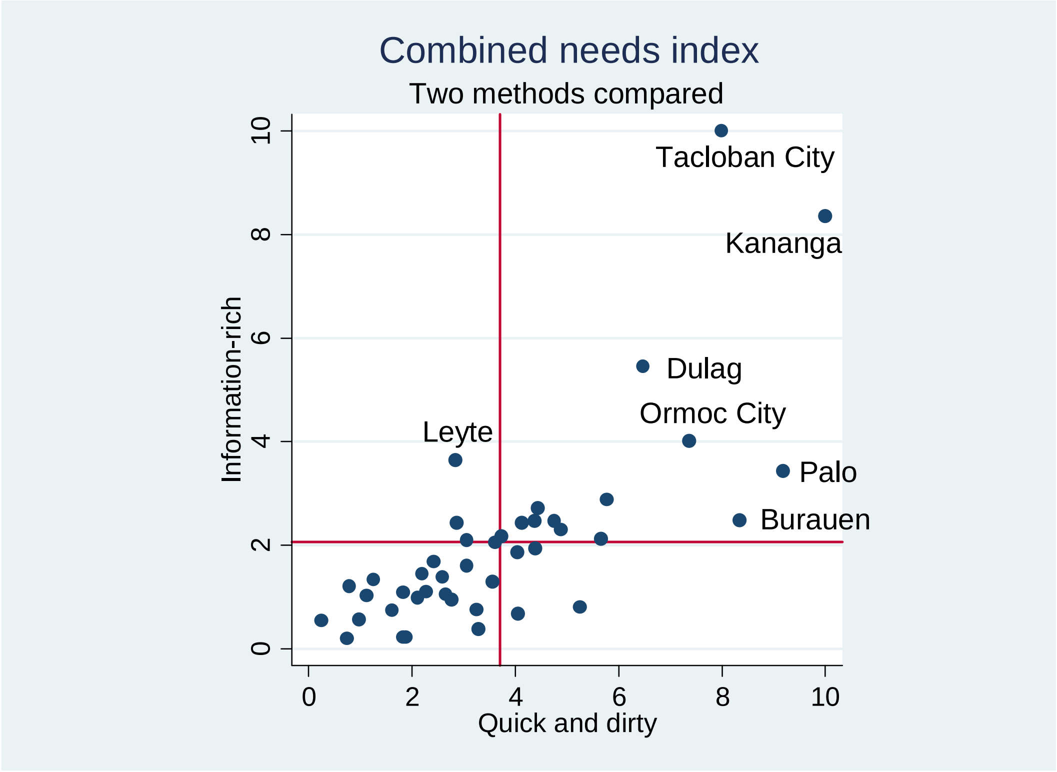
The index is useful to a modest degree only. It demonstrates that in a small proportion of all municipalities (2 out of 43) the aggregate needs are significantly larger than in each of the remaining communities. A ranking-based method would have underestimated these differences. But: Tacloban City and Kananga account for only (10 + 8.40) / (sum of scores of all 43 municipalities) = 18.4 / 88.6 = 21 percent of the aggregate needs. The index differentiates poorly among the municipalities claiming the other 79 percent.

Robust: Without formal testing, it may be assumed that Tacloban City's and Kananga's aggregate needs are significantly larger than those of any other affected municipality in the sample. The differences between any two of the rest will disappear rapidly as we factor in increasing levels of measurement error. This, of course, holds equally for ranking-based models, which create the illusion of significantly different large groups, when in fact the robustness of ranks is no greater than that of the underlying variables.

Is "quick-and-dirty" good enough?

Finally, this diagram shows the correlation between the scores of the quick-and-dirty vs. the information-rich methods (both re-scaled such that their maxima are 10).

Figure 13: Comparison of quick-and-dirty vs. information-rich



The product-moment correlation between the two measures is 0.72. If we accept the square of this figure (0.52) to be a rough statistic of their mutual information, we may think that the information-rich version, with its better balanced sub-indices, outperforms the quick and dirty. The latter considerably overestimates the needs of three municipalities - Ormoc City, Palo and Burauen, and underestimates Leyte's. However, it does agree with the information-rich method in signaling Tacloban City, Kananga and Dulag as high-needs municipalities. Similarly to this method, it places the majority in the comparatively low-needs group (lower-left quadrant).

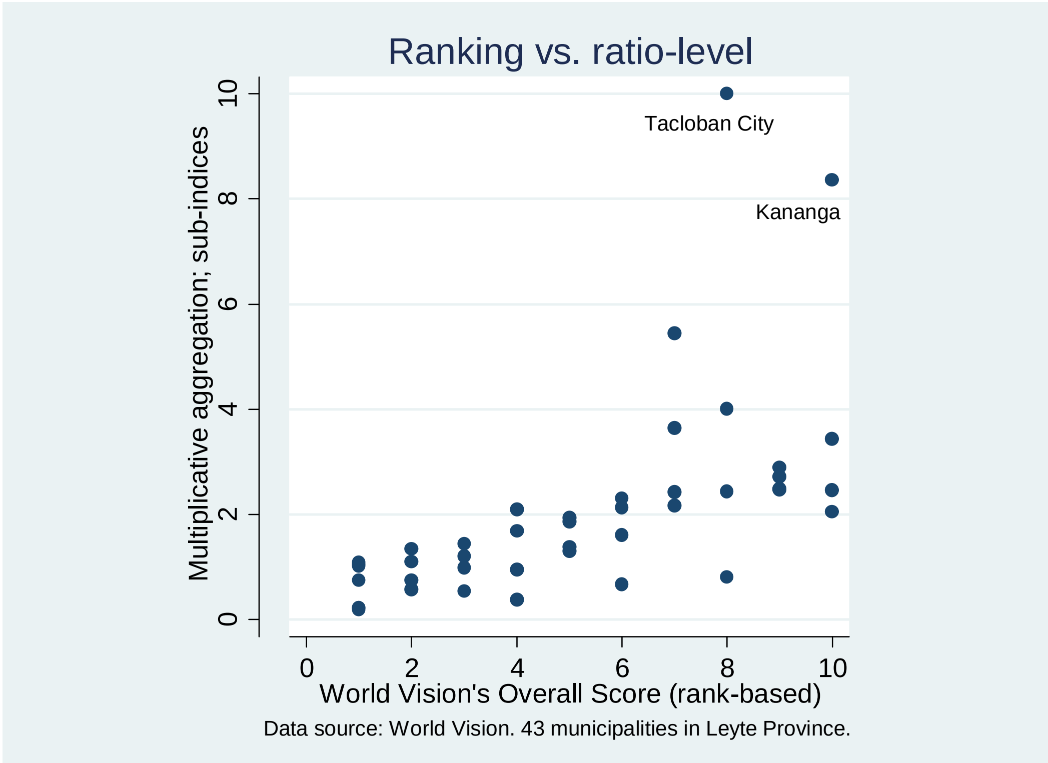
Information is costly. More information costs more to collect and keeps consumers waiting longer while they incur opportunity costs. The information-rich method depends on a statistical algorithm that most assessment teams cannot easily access. Without it, indicator weights may have dubious justification, or none.

Given costs and risks, the quick-and-dirty method seems preferable. If its results are not good enough, then the next consideration is to keep some of its components separate.

Ranking vs. ratio-level

Before we conclude this section, some readers may want to see how World Vision's original "Overall Score" compares with the needs index that we computed using separate sub-indices and multiplicative aggregation. World Vision's model involves weighted sectoral indices that are ranked, combined and re-ranked to the Overall Score. This score is rounded to integers between 1 and 10. The rounding is visible in this scatter plot in the dots aligned vertically. Our construct is a continuous variable.

Figure 14: Comparison of World Vision's and our alternative scores



A number of common as well as of divergent traits leap to the eye:

In common:

- The two municipalities that the alternative model identified as having far greater aggregate needs than the rest received high scores in the World Vision model as well.
- If we form two groups, one with Overall Scores 0 - 6, the other 7 - 10, these are fairly well discriminated in the alternative model.

Divergent:

- The ratio-level method assigns a modest score (< 5) to most of the municipalities that World Vision scored high.
- In the lower group (World Vision's 1 - 6), the ratio-level method recognizes small differences only. All the alternative scores are smaller than 2.5

In the eyes of the ratio-level index design, most in World Vision's upper group (7 - 10) are false positives - units with exaggerated scores. In the eyes of ranking believers, the ratio-level scaling creates too many false negatives - units with inappropriately low scores. Ultimately, this will come down to a question of the "policy scale". Do the assessment consumers prefer a smooth distribution of scores? Are they willing to work with asymmetry and discontinuity that focus attention on a few units?

The differences that the World Vision (and other) models make in the low range (here of the Overall Scores between 1 and 6) are largely illusory, regardless of policy. They are unlikely to be robust to reasonable assumptions of measurement error or changes in indicator weights.

The comparison also suggests that most, if not all, methods, agree on *some* of the extreme cases, at either end of the needs spectrum. But is that partial agreement good enough when in fact we have good indicators that discriminate well?

We now have enough material to pull the diverse thread together into some common challenges - challenges that are likely to recur in many rapid needs assessments.

Common challenges

The review of the four prioritization matrices revealed a number of problems that are likely to recur in other assessment settings as well. Every disaster, however, is different; findings therefore need to be formulated with the necessary flexibility, as agenda points rather than prescriptions.

The need for explicit models

Prioritization matrices are measurement models. It is not clear what they are intended to measure. What is the underlying key concept? "Disaster impact" needs to be explicated, but with what specifically? Consider these possible interpretations: Anticipated *excess mortality* if no relief arrives? Current *relative deprivation* compared to pre-disaster basic needs fulfillment? Anticipated *rehabilitation needs*? None of these is mentioned in the prioritization matrices; yet all can be meaningful expressions of the disaster impact.

As a stop-gap, we use the idea of *"unmet needs"* in the wake of the disaster. This is a global needs concept, which will be broken down as soon as sectoral plans take shape. Meanwhile it allows us to formulate a substantive model, however primitive, as developed in the previous section. One of its benefits is to sharply distinguish between magnitude, impact and pre-existing conditions.

It is the dialogue between process and measurement models that will advance the needs assessment discipline, as the tension between theoretical and experimental physics has done in the natural sciences. At present, process models appear to be absent. They need to be built.

Indicator-level criteria

Acquisition and use of information in needs assessments are opportunistic. Some information environments are rich, some are sparse. Quality varies, among types (variables) as well as among observations (records). Quality results from positive values on several criteria. Many of these are repeated in research design, information management, etc. textbooks and are hardly new. Most of these prescriptions are taught in environments and for purposes that know less time pressure than needs assessments after disasters must face. Under time pressure, the need for compromise is stronger. What follows is a cookbook without recipes because rarely will all the desired ingredients be available. Nevertheless, these criteria are usefully considered in indictor selection.

Cost of the indicator

Direct cost of gathering, processing and disseminating the data needed
Opportunity cost by displacing some other desirable activity
Damaging the information when it should rather be used in a non-indicator format

Value of the indicator

The type and amount of uncertainty that the addition of the indicator to other information reduces in decision making.

The uncertainty reduction depends on several attributes of the indicator:

<u>Speed:</u> The earlier the indicator is ready in usefully complete shape
<u>Relevance:</u> The strength of association with some concept that we estimate in order to inform the decision
<u>Validity:</u> The indicator is strongly correlated with the underlying variable, which we cannot observe directly
<u>Certainty of the indicator itself:</u>

- Clarity: the ease of communicating meaning and scale
- Completeness: the proportion of valid observations
- Discrimination: the observed values do not cluster in a small subset only of the relevant value set
- Reliability: the plausible belief (rarely the evidence!) that repeated measurement will obtain the same values
- Accuracy: the closeness of the observed values to the true ones
- Precision: the degree of resolution on the reported scale
- Measurement level: higher levels (ratio, interval) permit stronger analytic operations than lower levels (ordinal, nominal)
- Recency: the time that lapses between when the reported event happened and when the indicator covering it is used
- Updatability: the chance that the values first recorded can be updated with more recent ones

Optimal redundancy (a property of the indicator set, not of an individual indicator)

- Substantive redundancy: The indicators represent sufficient topics within the range of concerns, with some, but not excessive overlap for bridge-building

Statistical redundancy: The indicators measuring the same concept are correlated, but not too strongly. They compensate for each other's measurement errors, yet each "covers a bit of new ground".

The magic triangle

In rapid needs assessments, indicators are developed in the *magic triangle of cost, speed and quality.* They evolve not only because more data become available over time, but also because the type of decisions that they (hopefully) will inform changes. The relevance of the indicator system itself changes, with more information going into calculations tied to resource variables, such as budgets and distribution reports.

The delay that World Vision experienced in its effort to include livelihood indicators in the prioritization matrix is a good example. Such data were being collected (some can be seen in the *barangay* sheets), and have most likely since been used in the fine-tuning of the response. But it may be irrelevant by now trying to work them up as an element of a prioritization matrix.

The magic triangle is the reason why, sometimes, "quick and dirty" is better.

How many indicators

The four matrices differ in the amounts of distinct types of information that they incorporated. They also differ in the grouping principles, such as by presumed importance (via the weighting) or by building towards sectoral information matrices.

Trivially, information has benefits as well as costs. To complicate matters, assessment experts may work in - geographically and culturally - separate hubs and offices. Everybody is under time pressure; no single worker has the full picture of the relevance, quality, completeness and estimated arrival of the indicator data being pieced together from field surveys, remote sensing and secondary sources.

Thus, the question of how many indicators is as much group-dynamic as technical. Few, clean, reasonably complete, relevant indicators are preferable to numerous ones of highly variable quality. Our experiment with a quick-and-dirty vs. an information-rich index tilted towards favoring the former, if not very compellingly.

Relevance, quality and completeness evolve and are clarified over time. Updating occurs repeatedly, in already incorporated indicators and by adding new ones. The candidates need to be reviewed on those criteria, and also for their substantive and statistical redundancy. Do any two of them basically measure the same thing? Do they denote distinct things but are highly correlated? In this case, do we seek confirmation, by retaining both with split weights? Or do we favor novelty, with lesser weight for the one with the smaller coefficient of variation?

The interesting case of the Protection Cluster matrix

The Protection Cluster matrix is worth recalling at this point. The matrix creatively adds indicators of special vulnerability; these are less informative than the more common continuous indicators such as building damage or the poverty rate. They are binary (presence of special government programs) or count variables of rare events (counts of incidents or of villages with special programs). Protection-wise these indicators are highly important. Technically, it is doubtful that they should be on the same footing with the other indicators. The vulnerable groups to whom they speak may need special attention outside the matrix. The matrix format may not work for them. This challenge is difficult, calling for more deliberation and research than this note can offer.

[Sidebar:] Constructing a special protection needs index

As a mere thought experiment, we constructed a special protection needs index from two traditional and five novel indicators. All of these had been used by the Protection Cluster.

We normalized the indicators each to sum to 1 and allocated the weights using the Betti-Verma double-weighting rule (see page 41). Betti-Verma rewards (red) or penalizes (blue) indicators with higher or lower weights. Redundancy, measured by higher average correlations with the other indicators, lowers the weight[15]. Discrimination, expressed in larger coefficients of variation (= standard deviation / mean), heightens it. The aggregation mode is additive.

[15] The formula is more intricate than the simple mean, but "average correlation" captures the technical idea of redundancy in this context fairly well.

Table 9: Weights of a potential protection needs index

Indicators	Mean correlation with the other indicators	Coefficient of variation	Weight in the protection needs index
"Traditional vulnerability"			
Ratio IDPs to population	-0.05	1.13	0.12
Poverty rate	0.08	0.40	0.01
"Novel protection indicators"			
Specific threats:			
NPA-related incidents	0.15	3.07	0.10
Threatened villages / neighborhoods	0.20	3.67	0.10
Presence of special programs:			
Villages / neigh. with indigenous pop. (CADT)	0.03	6.97	0.35
Villages / neigh. with PAMANA program	0.21	3.00	0.09
Presence of NAPC programs	-0.14	0.55	0.22
		Total weights	**1.00**

Note: All indicators were normalized to sum to 1. N= 408 municipalities. Conditional formatting is separate for the three columns because their ranges differ. Darker reds mark conditions that increase weights (negative correlations; higher coefficients of variation; for darker blues, the reverse holds.

Some of the weights surprise. The weight of the poverty rate practically disappears. This indicator should be removed, to be used for other purposes such as in combination with a disaster intensity sub-index. The CADT indicator commands a high weight; it is not redundant and has excellent discrimination.

Whether one should admit weights this different on substantive grounds is another question (the ones we computed result in an extreme distribution of the index scores). But these weights are data-driven, on relevant criteria (non-redundancy, discrimination). They demonstrate the potential of building meaningful indices of protection needs, using non-traditional indicators. A lot of work will still need to be done to design valid and robust ones. Alternatives outside the matrix format should be tried as well.

Scores and ranks

One of the strongest messages that this review of prioritization matrices drives home is:

Ranking of interval and ratio-level indicators is detrimental. Ranking in this context gratuitously destroys information. The resulting final scores mask differences that the assessment consumers should be able to see, but are not allowed to.

This situation is different from concepts that are ordinal in the first place. Here ranking is natural and unavoidable. It is also different from other research situations in which the relationship of the indicator with the concept of interest is unknown, except for the assumption that it is monotonously increasing or decreasing. But the indicators like the percentage of damaged houses or the poverty rate are in a more informative rapport with the (not directly observable) unmet needs or expected excess mortality, etc. Ranking methods that would be appropriate in fuzzy-measure models are not suitable in prioritization matrices.

One suspects that the pervasive ranking habit must have some intellectual ancestry in the work of the NATF or related bodies. If it does, we have not seen it referenced. A rationale has yet to appear.

In short: Don't rank.

Missing data and imputation

This problem appeared in the reviewed matrices to a small extent only. It can be massive in other contexts.

Generally speaking, indicators with a substantial rate of missing values are difficult to use. There are situations where missing values stand for the total or almost total absence of the phenomenon that the indicator measures directly. If this is plausible, then missing should be set to zero (or another appropriate default value). Purists will do this in a new variable, duly documented, so that the original situation can be reconstructed if needed.

In other situations, missing is truly missing, with the true value likely different from any default value if a meaningful default exists at all. Thresholds above which imputation is reasonable are difficult to prescribe. But one may ask: In any important indicator with currently fewer than 10 percent missing values, should we impute these to the median of the observed values, in a separate variable ready to be updated?

There are situations where substantively similar indicators are available, but their definitions vary across regions, depending on organizations and reporting formats. For some districts, the proportion of households enrolled in microfinance programs may be available, for others the sum of outstanding loan balances (and the population). Sanitation indicators may differ between safe water access in rural, and piped supplies in urban environments. Under favorable circumstances, one can roughly be translated to the other, via concepts like test equivalency. The translation makes the data more uncertain, but also more comparable.

Linking pre-crisis conditions and disaster impact

The basic intuition is that, other things being equal, the disaster harms those more who were more vulnerable to begin with. This is the rationale for including pre-existing conditions in the impact scoring. Two questions naturally arise: how should we measure those conditions? And: how should the measures be connected to the score?

The primary indicator used for this purpose in the Typhoon Yolanda matrices is the poverty rate, based on surveys taken a few years before (only the Protection Cluster matrix includes some other pre-existing conditions). Three of the four matrices weight and include the rate in a purely additive aggregation function. For the Protection Cluster, we were able to demonstrate that the weighted poverty rate accounts for a quarter of the final score variance.

The question is how in the logic of pre-existing conditions the aggregation should proceed. Our model assumes that the unmet needs increase with the poverty rate, other things being equal. This can be expressed as "in proportion with some monotonously increasing function of the poverty rate". If so, the poverty rate should be incorporated multiplicatively. For simplicity, we multiply by the untransformed poverty rate. If this is thought to give poverty excessive consideration, the rate can be dampened with an exponent smaller than one, such as 0.5 (the square root of the rate). This is a policy question. The exponent expresses the poverty orientation.

When multiplicative doesn't work

Not all indicators of pre-existing conditions lend themselves to multiplicative aggregation. Those with substantial zero readings could not be used in this way. They would have to be transformed, such as by $a*rate + c$, $c > 0$, but this would be arbitrary unless we have an informed guess of a and c from some other relevant study. If there are several relevant pre-existing condition indicators available - say, poverty and malnutrition rates -, they can be additively aggregated in a sub-index, using weights according to the Betti-Verma formula (see above).

The selection of indicators will always be opportunistic; there is usually no time for retrospective surveys (although later during recovery much effort may be needed to reconstruct pre-disaster states, such as land titles). Their validity is a technical concern; their relevance and weight a matter of policy.

Linking magnitude and intensity

The matrices mix magnitude and intensity indicators, seeking a balance in their influence on the final ranking. Most magnitude indicators are absolute counts of items of interest. They have not been divided by a population or event-count denominator (they may be normalized within their own distributions, but they are not rated to anything else). First and foremost among them is the number of affected persons. This is an indispensable indicator even if both definitional and counting challenges abound, and substantial revisions up and down are to be expected.

Of intensity indicators we tend to think that they must be rates or proportions, resulting from counts divided by some suitable denominators. As noted before, unrated indicators such as wind speed too can figure in this category. Some may even be called magnitudes (as in earthquakes) while they are useful as intensity measures. Population-wide numerators and denominators are not necessary to form intensity indicators; sample data, or eye witness first impressions, produce intensity estimates.

Nor do numerator and denominator need to have the same units. Some of the matrices multiplied the number of damaged houses by a constant household size before dividing by municipality population. This makes for intuitive interpretation, such as in "an estimated X percent of the population are homeless". When we normalize the indicator, it does not make a difference.

Indicators that are not clearly related to either magnitude or intensity

Indicators of pre-existing conditions can express either magnitude or intensity. The poverty rate used in the four matrices is an intensity measure (and, one should add immediately, it is only one of a class of poverty measures, besides the depth and severity of the poverty). Magnitude indicators can claim equal relevance, depending on the assessment challenge at hand. If a country had suffered another disaster a short time before, the absolute number of persons still displaced just before the most recent struck will be an indicator of interest.

For some indicators, it is hard to decide whether they are measures primarily of magnitude or of intensity. This is true, for example, of several of the indicators that the Protection Cluster introduced. The presence of special government programs that its matrix recorded signals underlying problems (e.g., violent conflict) which we believe complicate the disaster impact. Problems usually cross some threshold before a special government program sets up shop in a new locality - the presence of the program thus proxies for the intensity of the underlying problem. However, if the measure is the number of subunits in which the program is active (*barangay* within municipalities), this will somehow be in rough proportion to the population covered - and thus related more to magnitude. One can argue that for pre-existing conditions the difference between magnitude and intensity indicators matters less. Still, one has to be vigilant about their logical consistency.

Grouping indicators in sub-indices

These considerations matter in the measurement model, which should support the intrinsic model of disaster impact as closely as possible. If we accept that impact is a function of magnitude, intensity as well as of pre-existing conditions, such as in

*Needs = k * Magnitude * Intensity * f(Pre-existing conditions)*

then indicators should be grouped by magnitude, intensity and pre-existing conditions. For each group, a sub-index should be formed from its associated indicators. The final impact index will result from weighting and aggregating the sub-indices.

The reviewed matrices followed different ordering criteria. The Protection Cluster divided its indicators by importance, into primary, secondary and tertiary groups with decreasing weights. This followed assumptions of what information would be most critically needed in the response. World Vision pursued a sectoral orientation anticipating its information needs for specific response plans. UNOCHA's matrix has relatively few indicators; it did not form sub-indices.

The absence of a process model and of the alignment of the measurement model with it is regrettable. It remains unclear what the final scores in these matrices measure. World Vision's pragmatic orientation is laudable. Yet, just because its workbook is so neatly designed, it would have been easy to figure magnitude, intensity and poverty in clear distinction, all of them in one summary worksheet.

Presenting magnitude and intensity side by side

The assessment consumers should be able to see the magnitude and intensity of the disaster separately. This facility should be granted in addition to the measure of aggregate impact that combines magnitude, intensity and pre-existing conditions. There are several reasons for this.

- The government is sovereign in determining how many people should be considered affected by the disaster. This number defines the magnitude more than any other. As we have seen, it is liable to substantial revision in the days and weeks following the onset. It should not compromise the operational picture of the intensity of the disaster.
- Updates to the intensity estimates should follow their own logic. An intensity index need not necessarily include the proportion of people affected if this measure is considered too vague. Other indicators may substitute for it.
- Response planners need to know both absolute and relative figures. Intensity is closely linked to the urgency of humanitarian needs. Magnitude informs about the size of the publics to be served. Unless they see both, planners are bound to mix heterogeneous publics in the same projections, such as relatively mildly damaged large urban communities with more heavily impacted rural villages.
- The elaboration of a combined impact measure may take more time because models are being debated, or data on more indicators deemed critical are still awaited. Meanwhile, rudimentary magnitude and intensity estimates are available and should be used.

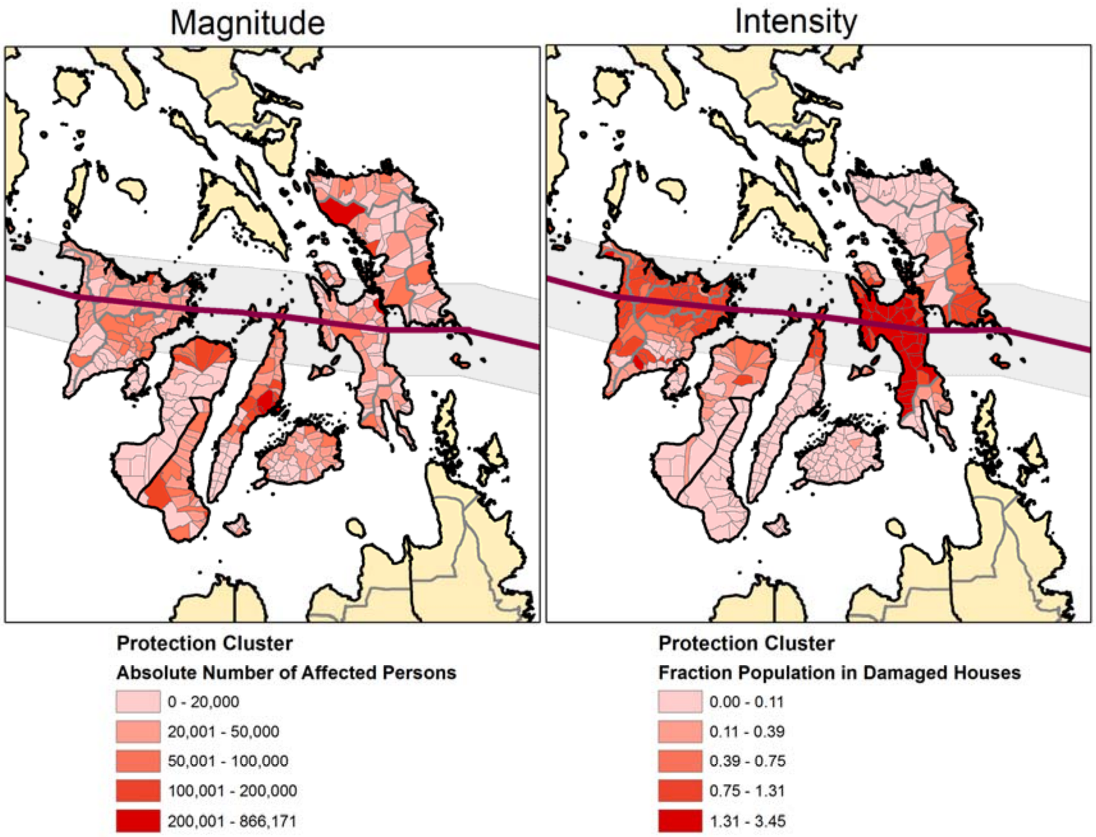
It should be possible to work out pragmatic compromises that respect the priorities and attention span of assessment consumers. Two maps, one representing the magnitudes of impact in local units, the other the intensity, can be shown side by side. In the main part of updated reports, shortlists of the most impacted areas can be printed, with complete lists relegated to the appendix or kept only in spreadsheets. The shortlist may display small sets of areas that have the highest values either on the magnitude, the intensity, or the pre-existing condition sub-indices. A bottom row gives the averages of all assessed areas. This or similar assessment products contrast extreme vs. typical areas as well as different drivers of impact. The perspective is dynamic, letting users expect updates that incorporate more precise figures and/or additional indicators.

The bottom line is that magnitude and intensity are distinct factors in the calculus of aggregate impact. Both information pieces have value for the response planning.

[Sidebar:] The geography of magnitude and intensity

These two panels display the distribution among municipalities ranked by the Protection Cluster. The left panel shows the absolute number of affected persons as the magnitude measure. The right panel relies on a housing damage ratio to express the intensity of the disaster impact.

Figure 15: Maps of disaster magnitude and intensity



The legend categories were formed by different methods. The category ranges for affected persons were successively escalated; the bounds are intuitive magnitudes. The ranges for intensity reflect natural breaks in the distribution. They exceed one because a uniform household size was used in the numerator, and the number of "affected persons" was the denominator of the intensity measure (calculated as damaged homes * 4.6 / affected persons).

Despite these ambiguities, differences between the two distributions leap to the eye. Some of the areas with high numbers of affected persons are far from the path of the typhoon. Unsurprisingly, magnitude is less strongly associated with the distance from the path than intensity is.

These different behaviors advocate for a nuanced presentation in assessment reports. While it is appropriate to combine magnitude and intensity measures (together with pre-existing conditions) in a combined index of needs, users should be able to evaluate the distributions of magnitude and of intensity separately, side by side.

Robustness to measurement error

Measurement error is pervasive in social research and is particularly large in rapid needs assessments after disasters, given the turbulence and uncertainty in the task environment.

Many initial findings have to be modified once more settled conditions permit re-assessments that measure more accurately. At the same time, it stands to reason that the initial pattern of starkly different impacts will persist to a significant degree if and when the impacts are re-assessed. A community in which 100 percent of all buildings were reported destroyed in the first assessment is unlikely to be rated better-off than one with a guesstimated 50 percent loss when a more detailed assessment is conducted.

The challenge is to make reasonable assumptions about levels of measurement error and to evaluate how robust the scores and ranks are under these assumptions. The demands of time and of modeling expertise may not warrant formal testing. But at least on the outliers, some crude testing can be done quickly. Questions like "*Would the unit with the highest impact score still be in the top quintile if the true value of indicator X were only half of the reported?"* are legitimate and easy to answer in a spreadsheet. As assessment coordinators become more familiar with the disaster environment and the data collected on it, ideas on the reliability of the various indicators will firm up. They will inspire more pertinent questions and better informed checks.

[Sidebar:] Robustness at different error levels

We tested a simple quick-and-dirty index of unmet needs for robustness. Using Protection Cluster data, we computed the index as the product of the number of affected people (for magnitude), the proportion of totally destroyed buildings (for intensity), and the poverty rate (for pre-existing conditions).

We varied the levels of mean measurement error in affected people and the destruction rate. For didactic simplicity, we assumed that poverty had been measured without error. We assumed also that the errors in the two indicators were not correlated. This seems problematic but we have no basis to assess the strength of the error correlation and are more interested in varying the extent of the two errors.

The table presents the mean standard deviation in the ranks of the 408 affected municipalities in response to combinations of different error levels. These standard deviations are a good expression of the uncertainty created by measurement errors. The absolute figures mean little[16]. What matters is the relative increase as we step up the amount of error in the two indicators.

[16] Some readers may expect the mean absolute deviation (MAD) in ranks as a more intuitive measure of the uncertainty. For normal distributions, the MAD is expected to be sqrt(2 / pi) = 0.798 of the SD (Wikipedia 2014a). However, ranks are bounded, and at the extremes of the index distribution, the ratio MAD/SD may be farther from that value. In either case, we are interested in the relative movement.

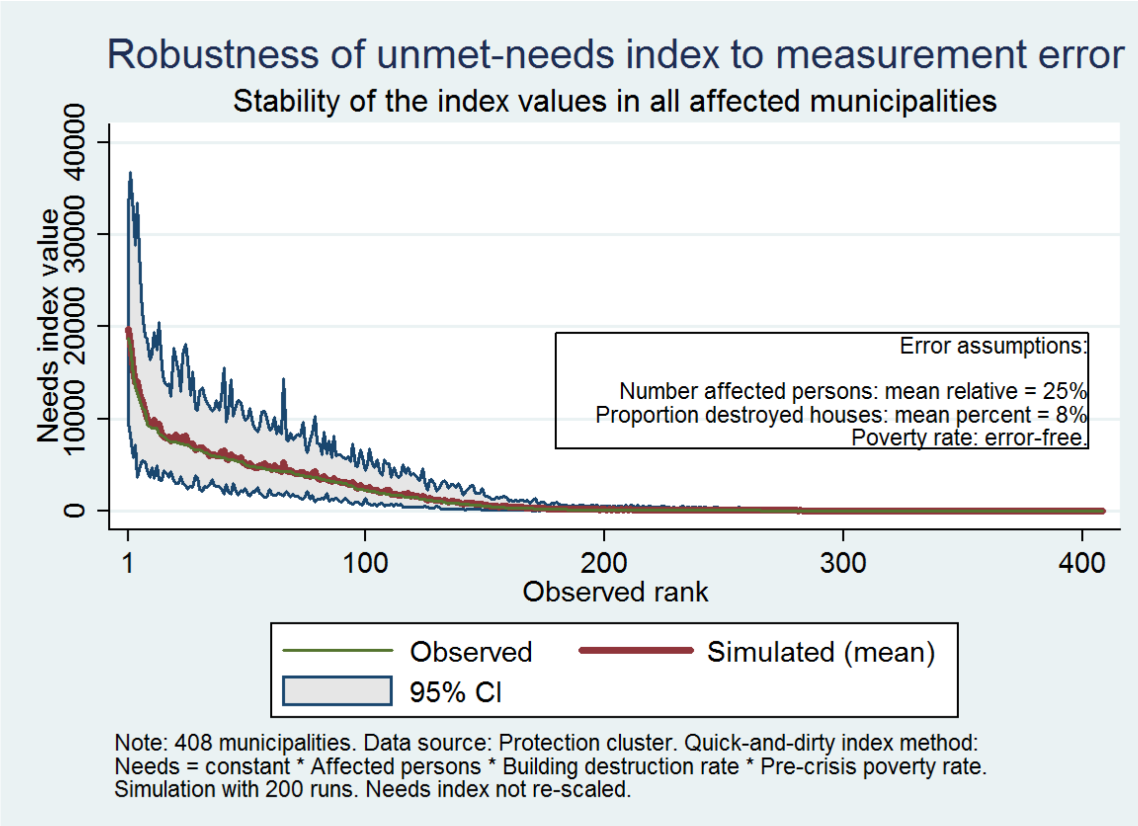
Table 10: Changes in rank in response to measurement error

Index of unmet needs — Mean SD of the change in ranks under measurement error								
		Building destruction						
Affected persons		Step	0	1	2	3	4	
Step	Mean relative error in percent:	Mean error in percent:	0.0	8.0	18.2	29.7	41.6	**Overall**
0	0.0		0	23	40	54	65	**36**
1	24.7		15	27	42	55	66	**41**
2	54.1		27	34	46	58	68	**46**
3	94.8		37	41	51	61	70	**52**
4	157.6		46	49	57	65	72	**58**
Overall			**25**	**35**	**47**	**59**	**68**	**47**

As we can see, the error behavior is not explosive. At step combination 1-1, the rank change SD is approximately 27. At 4-4, it is 72. This is less than a tripling although the error levels rose more than five times. Even in the extreme case - if the municipalities on average changed the ranks by 72 when re-assessed -, most of them would be found in the same or in a neighboring quintile of their initial ranks (because 72 < 408 / 5).

This indicates relative robustness. We can show this also in the behavior of the index scores. For this graph, we assume that the step combination 1-1 reflects a realistic assumption on error levels: 25 percent mean error in affected persons, 8 percent in the proportion of destroyed buildings. The first is set higher because of the difficulty to account for displaced persons. Building sites are static and thus - presumably - easier to evaluate for destruction.

Figure 16: Robustness of the unmet-needs index



The graph shows the 408 assessed municipalities ranked by the observed values of the quick-and-dirty index. The values were recalculated with simulated errors 200 times. The mean of the simulated values follows the observed ones closely (red and green lines). Their variation, however, is significant. Most of the simulated values - 190 out of 200 - fall between the blue lines bordering the grey-shaded area. The simulated mean and the confidence bounds are spiky because the relative contributions from the three indicators vary from municipality to municipality.

Nevertheless the index is robust. To see this, let us look at the top-ranked case (#1 = Guiuan Municipality in Eastern Samar Province). Its observed needs index value was 18,796; the mean simulated value is 19,727. The confidence interval for this is wide, from 9,577 to 32,866. Nevertheless, we see that the lower CI value is higher than the upper CI value of almost all other municipalities with ranks 70 - 408. Thus, with good certainty, Guiuan is in the top quintile of municipalities as regards their unmet needs. In fact, it can be shown that, at this level of measurement error, none of the twenty top-ranked municipalities is likely to rank lower than 80 if the true scores were known. This should lead us to consider that the needs index is robust. Given the nature of the disaster, with its steep intensity gradient, this is hardly surprising.

Recommendations

The section on common challenges gave a fair amount of detail for each of them. The recommendations, therefore, can be stated succinctly. They are limited to the use of indicators and do not delve into other aspects of good practice. ACAPS offers a number of notes on data collection, management and analysis; this guidance applies also to the treatment of the information that feeds indicator-based systems.

These recommendations simulate a step-by-step process from the definition of what is to be measured to the distribution of the disaster affected units on a final measure.

Draw a process model

Define the nature of the final object that you want to measure (e.g., "basic needs gap"). In one equation, write down 2 - 4 basic components and their basic relationships with the final object (e.g., "Gap = Population * Physical destruction * Previous social exclusion").

*The process model must remain simple, ideally diagrammed out "on the back of an envelope". Defining the components on both sides of the equation is essential, as is their shared understanding among the key players. Suppose there is no consensus on the feasibility to measure physical destruction in useful time. You might then replace it with some meaningful concept for which measures are ready: "Basic needs gap = Population * Exposure to destructive force * Previous social exclusion".*

Draw a measurement model

List information types that plausibly speak to each basic component. List variables in actual data sets that can be acquired within the constraints of speed, cost and quality. Anticipate the likely compromises and evaluate whether the basic component is likely to be measured satisfactorily. If unlikely, revise the process model.

Assume again that "physical destruction" cannot be measured in useful time. But "exposure to destructive force" can. You measurement model for exposure could be:

Exposure = 1 /

*((distance from the storm path to centroid [in km] + 10) *
(lowest altitude in the municipality [in meters] + 5) *
(shortest distance from the coastline [in km] + 2))*

conceiving of high exposure as the synergy of three necessary conditions (close to the storm path AND low-lying land AND coastal. This implies a simple model of wind force and storm surge, with reasonable constants added to regulate asymptotic behavior (avoiding division by zero; e.g. the storm surge is no higher than 5 meters). These components can all be computed remotely, using 3D GIS models.

Acquire the data and evaluate them variable by variable

In as many iterations as needed with the gradual acquisition of more data, evaluate each variable being filled for fitness as an indicator:

- Study the univariate distributions through histograms and, if maps are feasible at this point, in space.
- Assess low and high outliers, missing values, geographic isolates for their plausibility and impact on the measurement model.
- Do the same for the transformed variables that will likely be needed, chiefly ratios to population or to pre-existing assets.

Interpret the findings from inspecting each variable in the light of the refined quality criteria enumerated on page 51:

- Relevance
- Validity
- Certainty, which can be broken down into:
    - Clarity
    - Completeness
    - Discrimination
    - Reliability
    - Accuracy
    - Precision
    - Measurement level
    - Recency
    - Updatability

Determine for each variable of interest and for their transforms whether it is currently fit to be an indicator, a candidate awaiting improved data, or to be preserved for other potential uses.

Assume that you measure pre-existing conditions exclusively through the most recent poverty rates, and its estimates are missing for roughly one in every ten municipalities. Should you give up on modeling pre-existing conditions or on scoring those with missing values? No. Instead find a way to "impute" (= assign reasonable) values.

Depending on how much time you have, different imputation strategies may work. Regardless, make a copy of the original variable and call it something like "PovertyRate_Imputed". In the simplest approach, replace missing values with the median of the observed in each province. If you have more time to tinker with the data, map the poverty rate. In rural areas, replace missing values with the median value among adjacent municipalities. Urban areas are trickier because poverty rates may vary more pronouncedly over short distances. Use the median of adjacent rural units and *adjust by the national urban - rural differential, which most national poverty studies will report. The imputed values are prime candidates for updating once the areas can be visited and local key informants can be interviewed.*

Decide between "quick-and-dirty" and information-rich options

Having assessed the information extant, decide whether to aim for a quick-and-dirty impact measure or for one richer in information.

Prefer **quick-and-dirty** if, on balance, these considerations prevail:

- "Time-to-market" is very important
- For each component of the process model, you have one indicator of satisfying quality
- Sub-indices are not feasible, because of the dearth of good indicators or the difficulty to determine weights or likely incomprehension on the part of assessment consumers
- Weights cannot be determined because of disagreement (policy) or lack of statistical expertise (data-driven)
- Assessment consumers accept that your first "quick-and-dirty" will be overwritten by one or more subsequent refined assessments.

Prefer an **information-richer** option in the opposite case and if you feel that it adds value beyond what simpler variants offer.

Assume that you already have assessed the municipalities on a quick-and-dirty measure that serves response planners as a good starting point. The collection of data on additional indicators is promising, but will take some more time. Should you work towards an information-richer variant of the impact measure?

This depends on whether, at this stage of the response, such a refined measure is still useful. If the wave of sectoral assessments is already rolling, an overall expression of unmet needs may no longer be keenly needed by the response community. Instead, the indicator collection may provide valuable context information for sectoral planners.

The following recommendation is mainly, but not exclusively, written for variants that use sub-indices. Normalization and outlier control apply to all.

For each basic component, build a sub-index

Normalize each indicator by dividing by its sum (so that each will sum to 1). Do not rank them.

Assign each indicator to a basic component.

For each component, produce a correlation table of its indicators. If available, produce a matrix plot (a graph combining the scatterplots of all pairs of indicators). Evaluate the indicators for outliers, very high and near-zero correlations, and for significant non-linearity (U- or M-shaped associations).

Mark outliers for plausibility tests. Winsorize implausible outliers (i.e., replace high outliers with the highest plausible *observed* value, low outliers with the lowest plausible observed); do not trim the sample.

Mark pairs of highly correlated variables for redundancy. Eliminate the less compelling indicator or mark both for lower weights (see below).

Re-examine an indicator that has very low or zero-correlation with all the others in its group. If it is not plausible that the indicator plays an important role for the basic concept, eliminate it from the group.

For indicators in significant U- or M-shaped relationships with other indicators in the group, try finding an explanation in terms of special effects of regions, social groups, or the history of the disaster. If unexplained, reconsider it for its substantive importance. If important, keep it tentatively and mark the particular association as needing explanation in the assessment report.

Consider the substantively appropriate aggregation mode for combining these indicators in a sub-index. For sub-indices, in most conceivable cases it will be additive. If additive, determine weights. Set the weights on policy grounds or data-driven (with the possibility of overwriting some on policy grounds if good enough reasons exist). For data-driven weights, use the Betti-Verma formula or a simplified manual version that assigns weights in proportion to each indicator's [coefficient of variation / (1 + mean correlation with the others in the group)] to reward rarer items and penalize redundant ones. Re-scale the weights so that they sum to one.

Compute the sub-index on the basis of normalized indicator values and the chosen weights. Evaluate the result through histograms and maps.

Repeat the process for all basic components. Remember that it may be perfectly all right to have one indicator per component only, certainly in "quick-and-dirty" models.

Assume that you want to combine these three indicators in an intensity sub-index:

- *Ratio of affected persons to pre-crisis population*
- *Ratio of IDPs to pre-crisis population*
- *Ratio population in damaged houses (based on a damaged building count and an assumed mean household size) to pre-crisis population*

You find these correlation coefficients:

	Affected	IDPs	Damaged houses
Affected	1.0000		
IDPs	0.2747	1.0000	
Damaged h.	0.0635	0.7995	1.0000

Without the possibility of further statistical analysis, you reckon that the second and third indicators can stand in for each other (their correlation coefficient is a high +0.80). You believe that building damage was measured more reliably than the IDP counts. Consequently, you drop the IDP indicator and form a subindex using the normalized versions of the affected and damaged housing ratios, with weights of 0.5 on both.

Combine the sub-indices in the final index

Determine the aggregation mode so that it makes sense in the process model.

If the sub-indices have many zero values, a multiplicative model may not be appropriate. Consider adding a (small) constant to the sub-index with many zero values, but if the resulting minima have no substantive plausibility (i.e., there is no way to decide whether to add, e.g., 1%, 10% or 25% of the median value), reject the multiplicative model.

Decide the weights or the sub-indices. The weights at this stage should be set on policy grounds, not data-driven.

For the additive model, rescale the sub-indices such that each sums to one. Re-scale the weights so they sum to 1.

The multiplicative model does not need normalization or weights for commensuration. However, in analogy to weights, consider dampening a sub-index (e.g. the poverty rate) with an exponent < 1 or amplifying it with one > 1. The exponents must reflect policy orientation (e.g. the importance given to pre-existing conditions).

Calculate the final index. For cosmetic reasons, you may rescale the index proportionately, so that its maximum is 10, 100 or whatever plausible end point. Do not rescale the minimum to zero or one (this would amount to a disproportionate transformation).

If you must rank, now, and only now, you may. Regardless, assessment ethics demands that you show the consumer the distribution of the untransformed final index.

Assume that your process model recognizes the protection of tribal populations as an acute need after the disaster. In the process model, you initially add protection as a separate component, as in

*Needs = Population * Physical destruction * Previous social exclusion * Special risks for tribal populations.*

When you consider measuring these risks, you realize that the only readily accessible indicators are: 1. the presence of special government programs for tribal populations, 2. the rate of deforestation in the past ten years, as mapped by a forest conservancy group.

This would result in a sub-index with numerous zero values (for areas without such programs and for urban areas without forests). In a multiplicative model of the overall unmet needs, these would lead to invalid zero need scores.

*You therefore revisit the process model and replace " Previous social exclusion * Special risks for tribal populations" by the unspecific "Pre-existing conditions". On the measurement side, you choose an <u>additive</u> aggregation of the (normalized) poverty rate and tribal protection sub-index, with weights that are politically acceptable.*

Critique, document and share the index and /or its components

Inspect the distribution of the index in histograms and maps.

Evaluate its distribution against the default expectation that disasters cause a small part of the affected area / population to have high values on the dimension that the index measures. A larger part is expected to have mid-range values, the majority to have low values.

Explain deviations from the default as the outcome of the known dynamic of this disaster (e.g. the vulnerability distribution) or as a failure of the index model, which therefore needs to be replaced with a better model.

Document how the model was designed, explain key parameters (particularly the weights and their policy or technical justification), and make a plan for updating or revision.

Publicize the index in a shape that the assessment users understand, such as in a clean prioritization matrix. Decide whether it is appropriate to present basic components as well (e.g., a map of the magnitude side by side with a map of the intensity of the disaster).

Consider this fictitious example: For a quick-and-dirty impact measure, you specify this simple process model:

*Unmet needs = Exposed population * Pre-existing conditions.*

You measure exposure as "Population / (Distance from the storm path [in km] + 10)", and pre-existing conditions as the latest poverty rate.

Assume further that the histogram of the resulting unmet-needs measure is bimodal. The distribution is not as expected. Rather, you find a substantial number of highly-impacted municipalities, relatively few moderately impacted ones, and the usual majority of low- or no-impacted ones.

Such a deviant pattern would have to be explained. For example, it could be caused by the fact that the areas closer to the storm path happen to be more densely populated.

If a rational explanation cannot be found in the distribution of the data, the model may be inappropriate. For example, the distance decay function "1 / (Distance from the storm path [in km] + 10)" may give too much weight to areas farther away; try one that decays faster such as "1 / (Distance from the storm path [in km] + 5)^2" ["^2" means squared].

*If you are uncomfortable with the arbitrariness of the exposure part, you may want to reject the model altogether. You may wait for the kind of data to come from the field that are required to calculate a more plausible "Unmet needs = Magnitude * Intensity * Pre-existing conditions" model. Of course, there is no guarantee that this measure will meet the default expectation. Assessment models themselves are inherently risky.*

Outlook

By 16 April 2014, five months after the typhoon, the Assessment Registry maintained at the "Humanitarian Response, Philippines" site[17] listed 212 assessments, of which 174 had been completed. Many of these may have been part of larger multi-location exercises, and to this extent 212 exaggerates the number of distinct assessment endeavors. Still, it is obvious that the four prioritization matrices that this note has reviewed were a small part only of the methodological variety employed in the total assessment effort. Most of it has remained outside our purview.

A humble view of our work, therefore, is in place. At best, we can say that the prioritization tools that we analyzed are elements in an evolutionary pool in which bits and pieces are varied, selected and retained as chance and fitness dictate. We have witnessed little of the variation directly, have noticed some selective pressures (all four opted to rank indicators) and have no means of predicting which of the selections made in the Typhoon Yolanda theater will be retained in assessments in future crises.

Our recommendations in the preceding section addressed the practical steps of processing indicators and indices. Looking to the future of prioritization methodologies, we can add some more abstract considerations. These again can be formulated in evolutionary terms of variation and selective retention (Campbell 1960; Burns and Dietz 1992).

[17] https://philippines.humanitarianresponse.info/assessment-registry.

Variation, in a prescriptive sense, means the need to do more experiments. We believe that experimentation in the following areas will promote the development of prioritization methods in humanitarian crises:

- Build conceptual models of what we want to measure, and of how we measure it. We captured this by the distinction of **process and measurement models**. Our illustration with an index of unmet needs offers one of many conceivable specifications. More should be proposed and debated, both for the overall concept (e.g., unmet needs) and for its key concepts (e.g., pre-existing conditions - magnitude - intensity). This should be a general debate, detached from any one particular emergency.
- Try **quick-and-dirty** methods, such as the one proposed to work with just three Yolanda indicators. Evaluate their speed, cost and information value against those of **information-richer** models. Are the latter feasible and productive in the short period between an initial quick-and-dirty assessment and the onset of sector-specific assessments?

Selective retention, prescriptively, relates to those elements of prioritization that should be repeated in future needs assessment, in suitable adaptation to other contexts. Conversely, some other practices should be dropped because they are sub-optimal or downright detrimental:

- Whatever the debates in the wider humanitarian community in general, when confronted with a real emergency, determine **what it is that you want to measure**, draw a small model of its major components, and define their connections to the key concept. Then consider how to measure the components, and how to adjust both the process and measurement models to the (changing) available information.
- **For greater learning** effect, document the rationales for important choices and drop practices for which none exists. Welcome discontinuities and outliers in index distributions because we learn more from the edges of experience than from averages. If outliers on key indices reflect real-world extremes, allocate more assessment resources to them, to gauge the true extent of impact and need.
- **Continue the engagement with decision sciences** and social indicator research from which the current prioritization matrix format emerged, in such directions as data-driven weighting (e.g., via the Betti-Verma formula), multiplicative aggregation and possibly other promising conceptual developments.

Each of the four matrices contributed something special to the evolutionary pool - from the rapid use of early public information, through the dual use of indicators as magnitude and as intensity measures, through the connection to response planning, to experiments with unorthodox indicators such as of protection needs. This wealth of variants lets us expect that prioritization matrices will absorb novel elements as their practice widens in future needs assessments. Assuming effective selection and retention mechanisms, they will continue to evolve as useful tools of humanitarian decision-making.

Statistical appendix

Figures and tables in the index are not captioned. The output from the statistical application STATA (Stata Corporation 2011) is unedited.

An Excel demonstration workbook is available for download from the same ACAPS Web page. Currently the file name is *Acaps_140527_Philippines_DemoDatatset.xlsx*. The data are a subset of the Protection Cluster prioritization matrix, with sub-index and index variables computed by us. One of the worksheets demonstrates an approximation, easily computed in Excel, of Betti-Verma weights (see below).

The Betti-Verma formula in index construction

The rationale for minimizing overlap

The sub-indices that correspond to the components of our process model of unmet needs resulted from *additive* aggregations of the concerned indicators. The weights were determined by a formula known as the Betti-Verma double-weighting rule. Betti-Verma **minimizes overlap** between indicators and **rewards well-discriminating** indicators with higher weights.

The technicalities are presented, to some extent, below. More importantly users need to understand why they would want to compute indices using this formula. This understanding is all the more important because textbooks often advise statistical procedures that achieve the very opposite - maximizing overlap.

Specifically, factor analysis is used to reduce the correlation pattern among candidate indicators to a small number of common factors. The index values are then computed as the scores on the first factor. Possibly several independent indices may be formed, each based on one factor, assuming the factors have a plausible substantive interpretation. Similarly, principal components seek to map a set of (unstandardized) indicators onto a number of independent components, such that the first exhausts the largest portion of the total variance. Both procedures, although with different objectives, emphasize commonality among indicators, in other words maximize overlap.

Whether an index should maximize or rather minimize overlap in its constituent indicators depends on conceptual interest. If the intent is to find one common factor that *causes* the various phenomena captured by the indicators, then we *maximize overlap*, i.e. minimize variation. Generally, concepts like ability, intent, strategy, wealth call for this approach. Thus, the wealth levels of the households in a survey sample may be unobservable, but the reasoning is that greater wealth is the enabling cause of the presence of more durable consumer goods. These we can observe. Procedures that estimate scores for the underlying wealth variable on the basis of recorded goods maximize the overlap. Rich households own expensive cars and jewelry; poor ones have neither.

By contrast, when we believe that the concept of interest is the *result* of the several phenomena captured by the indicators, we will want to *minimize overlap*. This is typically the case of deprivation studies that assume that it is the cumulative disadvantages of various kinds that ultimately lead to changes in the overarching variable of interest, such as life expectancy. This obviously is the assumption underpinning also the prioritization matrices used by the humanitarian community. There are multiple factors - we singled out magnitude and intensity of the disaster as well as pre-existing conditions - that combine to produce impact, needs, etc. Within each component, e.g. the magnitude, several sub-processes are at work that amplify or reduce the component's level, such as population displacement and physical destruction. The appropriate approach is to let them contribute broadly, but only to the extent that they don't overlap. Betti-Verma achieves this in the additive aggregation mode.

Technicalities

The Betti-Verma formula (Betti, Cheli et al. 2005) is implemented as one of the options in the user-defined STATA procedure *mdepriv* (Pi Alperin and Van Kerm 2009). In fact, mdepriv is so flexible that it lets the user decide whether to activate the entire double-weighting rule, or only the part that rewards statistical discrimination, or only the minimization of overlap. We have not made use of this subtlety and have employed both components together. Readers trying to implement the formula in Excel may find it easier to solve the discrimination part (which relies entirely on the indicators' coefficients of variation). The redundancy part of the formula is much more involved.

Pi Alperin and Van Kerm (2014, op.cit.: 2) present the two parts as follows:

Betti & Verma (1998) (and subsequently Betti *et al.*, 2008, Pi Alperin, 2007, 2008) adopted a more sophisticated double-weighting rule sensitive to both the relative frequency of items and the correlation among items. The correlation is taken into account so that two perfectly correlated items 'count as one' and only two uncorrelated items fully 'count as two'. To achieve this, Betti & Verma (1998) and Betti *et al.* (2008) defined item weights as the product of two components

$$\omega_j^{bv} \propto \left(\omega_j^a \times \omega_j^b\right)$$

with ω_j^a being the coefficient of variation of x_{ij} acting similarly to the frequency-based weights described above,[4]

$$\omega_j^a = \frac{\left(\sum_{i=1}^{N}(x_{ij} - \bar{x}_j)^2\right)^{1/2}}{\bar{x}_j N^{1/2}}$$

and

$$\omega_j^b = \left(1 + \sum_{m=1}^{M} \rho_{jm} I(\rho_{jm} < \rho_H)\right)^{-1} \left(\sum_{m=1}^{M} \rho_{jm} I(\rho_{jm} \geqslant \rho_H)\right)^{-1}$$

where ρ_{jm} is the correlation between items j and m and $I(\cdot)$ is an indicator function evaluating to 1 if the expression in brackets is true and 0 otherwise. ρ_H is a pre-determined cut-off correlation level.[5] ω_j^b is the inverse of a measure of 'average correlation' of item j with all the other items. The larger is the average correlation with item j, the lower is the resulting weight for item j.

In all cases, normalization to unity is achieved by setting

$$w_j = \frac{\omega_j}{\sum_{m=1}^{M} \omega_m}.$$

(reprinted with permission)

The cut-off level rhoH is set such that it "divides the ordered set of correlations at the point of the largest gap" (ibid., fn.5) unless the user opts for a different level.

As is obvious from the first bracketed term in the formula for the second part (*omega_b_j*), under extreme conditions, i.e. if *sum(rho_j_m * I(rho_j_m < 0)) < -1*, the weight for indicator j will turn negative. This can only happen if it is negatively correlated with some or all of the other indicators. This situation is atypical in a collection of deprivation-like indicators and would call for a review of the particular one.

What can the Excel user do?

1. Normalize each indicator by dividing it by its sum.

2. Depending on how many indicators the sub-index contains:

    a. If the sub-index is built on only one indicator, Betti-Verma does not apply (there is no weighting).
    b. With two indicators, only the first of the two rules apply (the second does not differentiate). Compute the coefficients of variation by dividing the

standard deviation by the mean and create the weights as $w_1 = CV_1 / (CV_1 + CV_2)$ and $w_2 = 1 - w_1$.

c. With more than two indicators, for the first rule, simply compute the coefficients of variation (normalization to unity can wait until after multiplication with the second rule). The big challenge is to find an approximation of the second rule formula is needed.

One conceivable approximation would require the user to make a correlation table for the indicators in point, using Excel's Data - Analysis - Data analysis - Correlation:

1. Create the correlation table.
2. For each indicator, compute the sum of the correlation coefficients with itself and with all others.
3. Take the reciprocals.

Then combine rule 1 and rule 2 results by multiplying the coefficients of variation with these reciprocals. Normalize such that these products sum to one. These normalized quantities are approximate Betti-Verma weights.

We have tested the approximation in Excel for the magnitude sub-index (see the demo workbook), with excellent results. To calculate the full unmet-needs index, we have used the facility provided by the statistical package STATA, through its procedure *mdepriv*.

The following section demonstrates the generation of an unmet-needs index. The weights of the sub-indices are determined by the Betti-Verma formula.

Generation of an unmet-needs index

on the basis of Cluster Protection data as well as imported malnutrition data

The index results from the multiplicative aggregation of three sub-indices. The sub-indices are computed by additive aggregation, with weights derived by the Betti-Verma formula. We also briefly discuss the handling of extreme values and visualization in histograms.

Creation of sub-indices

Magnitude

Definitions

```
              storage  display     value
variable name   type   format      label      variable label
------------------------------------------------------------------------------------------
affected        float  %8.0g                  Affected Persons
idpcumul        float  %8.0g                  IDPs (total, cumulative)
housetotal      int    %8.0g                  Houses totally damaged
htotalimput     int    %8.0g                  Houses totally damaged (imputed missing to zero)
housepartial    long   %8.0g                  Houses partially damaged
hpartialimput   long   %8.0g                  Houses partially damaged (imputed missing to zero)

Normalized:
magn_affected   float  %9.0g                  Affected persons (normalized to sum to 1)
magn_idpcumul   float  %9.0g                  IDPs (total, cumulative) (normalized to sum to 1)
magn_htotalim~t float  %9.0g                  Houses totally damaged (normalized to sum to 1)
magn_hpartial~t float  %9.0g                  Houses partially damaged (normalized to sum to 1)
```

Descriptive statistics

```
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
    affected |       408    33026.75    56399.26          0     866171
    idpcumul |       408    12610.14    21186.98          0     276509
  housetotal |       376    1430.258    2466.725          0      14132
 htotalimput |       408    1318.081    2398.864          0      14132
housepartial |       376     1482.16    3398.186          0      46553
hpartialim~t |       408    1365.912     3286.17          0      46553

-------------+--------------------------------------------------------
Normalized:
magn_affec~d |       408     .002451    .0041855          0   .0642803
magn_idpcu~l |       408     .002451     .004118          0   .0537439
magn_htota~t |       408     .002451    .0044607          0   .0262785
magn_hpart~t |       408     .002451    .0058967          0   .0835343
```

Compute weights for magnitude indicators and generate magnitude sub-index, re-scale with max = 100:

```
. mdepriv magn_*, method(bv) generate(magnitudescore)

Betti & Verma (1998) weighting scheme

Aggregate deprivation level:   0.0025

Deprivation level, weight and contribution to total, by item
 ------------------------------------------------------------------
                          |    Index    Weight    Contri     Share
 -------------------------+----------------------------------------
            magn_affected |   0.0025    0.3415    0.0008    0.3415
            magn_idpcumul |   0.0025    0.1663    0.0004    0.1663
         magn_htotalimput |   0.0025    0.2259    0.0006    0.2259
       magn_hpartialimput |   0.0025    0.2663    0.0007    0.2663
 -------------------------+----------------------------------------
                    Total |             1.0000    0.0025    1.0000
 ------------------------------------------------------------------
```

The *mdepriv* output lingo denotes as "Index" the means of the indicators, as "Contributions" the contributions that the weighted indicators make to the sub-index mean (also known as the "aggregate deprivation level", here 0.0025), and as "Shares" the contribution re-scaled to sum to 1. This terminology was borrowed from deprivation studies and is of little concern here.

Magnitude sub-index

```
               storage  display     value
variable name   type    format      label      variable label
------------------------------------------------------------------------------------------
magnitudescore  double %10.0g                  Magnitude score
magnitmax100    float  %9.0g                   Magnitude score (re-scaled with max = 100)

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
magnitudes~e |       408     .002451    .0035555          0   .0429998
magnitmax100 |       408     5.69998    8.268579          0        100
```

Intensity

Definitions

```
               storage  display     value
variable name   type    format      label      variable label
-------------------------------------------------------------------------------------------
ratioAffect     float  %8.0g                   Percent of Affected Person to Tot pop 2013
rimputAffect    float  %8.0g                   Ratio of Affected Persons to Tot Pop. 2013
                                                 (truncated to 1)
ratioIDP        float  %9.0g                   Ratio of total IDPs to population
ratioDamaged    float  %8.0g                   Ratio of Pop. with (Totally or Partially)
                                                 Damaged Houses to Affected Pop.
rimputDamaged   float  %8.0g                   Ratio of Pop. in Damaged Houses to Affected
                                                 Pop. (truncated to 1)
intens_rimput~t float  %9.0g                   Ratio affected to total pop 2013 (normalized
                                                 to sum to 1)
intens_ratioIDP float  %9.0g                   Ratio of total IDPs to pop 2013 (normalized
                                                 to sum to 1)
intens_rimput~d float  %9.0g                   Ratio pop. in damaged houses to total pop
                                                 (normalized to sum to 1)
```

Descriptive statistics

```
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 ratioAffect |       408    .7661142    .3918565          0    1.15214
rimputAffect |       408    .7404327    .3714224          0          1
    ratioIDP |       408    .3268645    .3689985          0   1.244828
ratioDamaged |       408    .4398802    .5584971          0    3.44577
rimputDama~d |       408    .3724832    .4183799          0          1
-------------+--------------------------------------------------------
Normalized:
intens_rim~t |       408     .002451    .0012295          0   .0033102
intens_rat~P |       408     .002451    .0027669          0   .0093343
intens_rim~d |       408     .002451     .002753          0   .0065801
```

Compute weights for intensity indicators and generate intensity sub-index, re-scale with max = 100:

```
. mdepriv intens_*, method(bv) generate(intensityscore)

Betti & Verma (1998) weighting scheme

Aggregate deprivation level:   0.0025

Deprivation level, weight and contribution to total, by item
 ------------------------------------------------------------------
                          |    Index    Weight    Contri     Share
 -------------------------+----------------------------------------
       intens_rimputAffect |   0.0025    0.2578    0.0006    0.2578
          intens_ratioIDP |   0.0025    0.3385    0.0008    0.3385
      intens_rimputDamaged |   0.0025    0.4037    0.0010    0.4037
 -------------------------+----------------------------------------
                    Total |             1.0000    0.0025    1.0000
 ------------------------------------------------------------------
```

Intensity sub-index

```
              storage  display     value
variable name   type   format      label      variable label
-----------------------------------------------------------------------------------------------
intensityscore  double %10.0g                 Intensity score
intensitymax100 float  %9.0g                  Intensity score (re-scaled with max = 100)

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
intensitys~e |       408     .002451    .0020211          0   .0065859
intensit~100 |       408    37.21543    30.68836          0        100
```

Pre-existing conditions

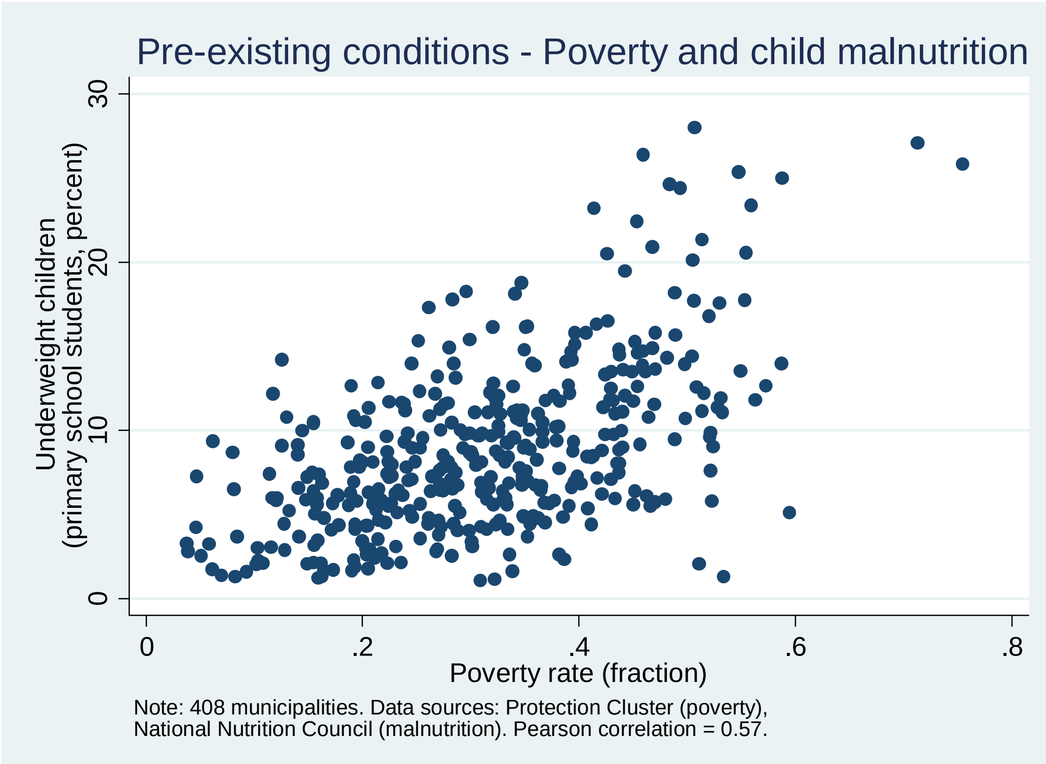
Definitions

```
              storage  display     value
variable name   type   format      label      variable label
--------------------------------------------------------------------------------------
povertyrate     float  %8.0g                 Poverty rate (as number poor / est. pop. 2013)
vuln_poverty    float  %9.0g                 Poverty rate (normalized to sum to 1)
TotalUW_Pc      double %10.0g                Malnutrition rate (percent, primary
                                               school students)
vuln_TotalUW_Pc float  %9.0g                 Malnutrition rate (primary school
                                               students) (normalized to sum to 1)
```

Descriptive statistics

```
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
 povertyrate |       408    .3153091    .1267646      .0374     .75461
vuln_poverty |       408     .002451    .0009854   .0002907   .0058658
  TotalUW_Pc |       402    8.893187    5.005149   1.060422   27.97619
vuln_To~W_Pc |       402    .0024876       .0014   .0002966   .0078254
```

Note the small difference in the numbers of observations, which results in small differences in the mean and in the contributions to the sub-index.



Compute weights for pre-existing conditions indicators and generate sub-index, re-scale with max = 100:

```
. mdepriv  vuln_TotalUW_Pc vuln_poverty , method(bv) generate(preexistscore)

Betti & Verma (1998) weighting scheme

Aggregate deprivation level:   0.0025

Deprivation level, weight and contribution to total, by item
 ------------------------------------------------------------------
                          |    Index     Weight     Contri     Share
 -------------------------+----------------------------------------
          vuln_TotalUW_Pc |   0.0025     0.5834     0.0015    0.5890
             vuln_poverty |   0.0024     0.4166     0.0010    0.4110
 -------------------------+----------------------------------------
                    Total |              1.0000     0.0025    1.0000
 ------------------------------------------------------------------
```

Pre-existing condition sub-index

```
              storage  display     value
variable name   type   format      label      variable label
-----------------------------------------------------------------------------------
preexistscore   double %10.0g                 Score of pre-existing conditions
preexistmax100  float  %9.0g                  Score of pre-existing conditions
                                                (re-scaled with max = 100)

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
preexistsc~e |       402    .0024641    .0010995   .0004483   .0067239
preexist~100 |       402    36.64731    16.35253   6.666938        100
```

Multiplicative aggregation

The sub-indices

```
              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------------
magnitmax100    float  %9.0g                  Magnitude score (re-scaled with max = 100)
intensitymax100 float  %9.0g                  Intensity score (re-scaled with max = 100)
preexistmax100  float  %9.0g                  Score of pre-existing conditions
                                                (re-scaled with max = 100)

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
magnitmax100 |       408     5.69998    8.268579          0        100
intensit~100 |       408    37.21543    30.68836          0        100
preexist~100 |       402    36.64731    16.35253   6.666938        100
```

are multiplied to form the unmet-needs index, which we re-scale with max = 100:

```
              storage  display     value
variable name   type   format      label      variable label
----------------------------------------------------------------------------------------------
UnmetNeedsMul~c float  %9.0g                  Index of unmet needs (multiplicative aggregation)
UnmetNeedsM~100 float  %9.0g                  Index of unmet needs (multiplicative aggregation)
                                                (re-scaled with max = 100)

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
UnmetNeeds~c |       402    12015.58    23012.41          0   294349.8
UnmetNee~100 |       402    4.082075    7.818049          0        100
```

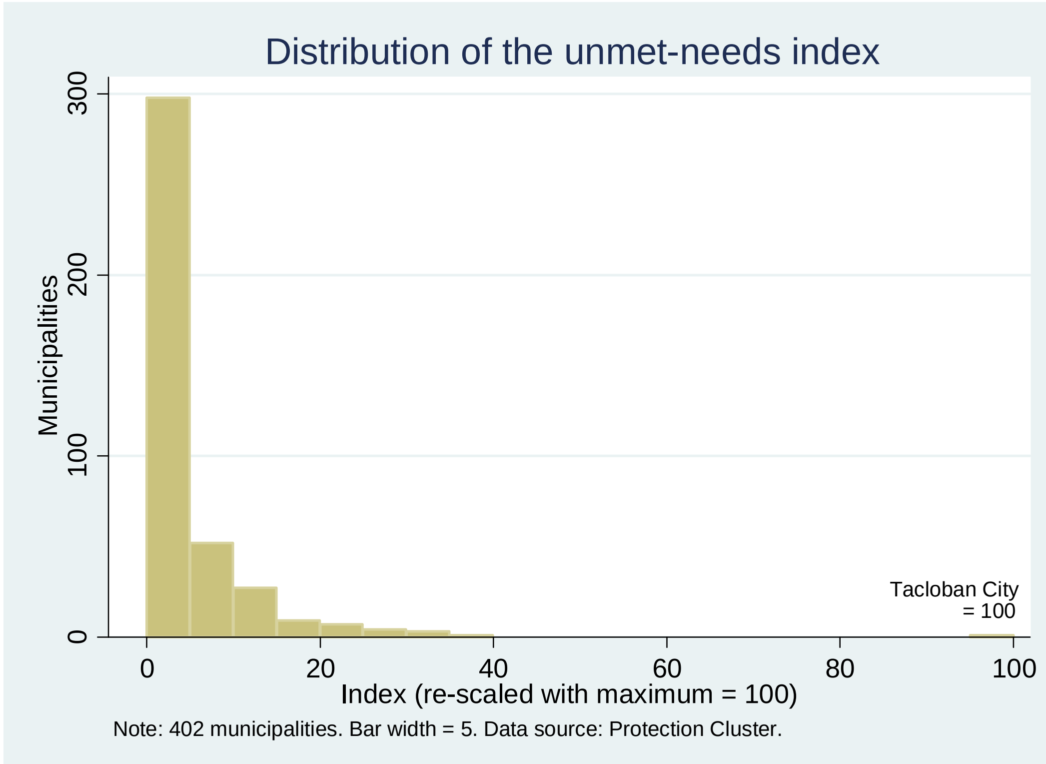
Outlier identification

The index has a highly skewed distribution and one extreme outlier. Sorting on UnmetNeedsMax100 descendingly, this is the list of the five municipalities with the highest index values:

```
. list province municipality magnitmax100 intensitymax100 preexistmax100 UnmetNeedsMax100 in 1/10

     +------------------------------------------------------------------------------------+
     |          province              municipality   magn~100   inte~100   pree~100   Unme~100 |
     |------------------------------------------------------------------------------------|
  1. |             LEYTE   TACLOBAN CITY (CAPITAL)        100        100   29.43498        100 |
  2. |             LEYTE               ORMOC CITY   51.12788   54.83699   37.40916   35.63244 |
  3. |            ILOILO                   CARLES   20.82018   84.75773   56.17022   33.67489 |
  4. |              CEBU             DAANBANTAYAN   28.74336   88.94603   38.26487   33.23538 |
  5. |             CAPIZ                    TAPAZ   17.25179   90.33174   60.55901   32.06192 |
     |------------------------------------------------------------------------------------|
```

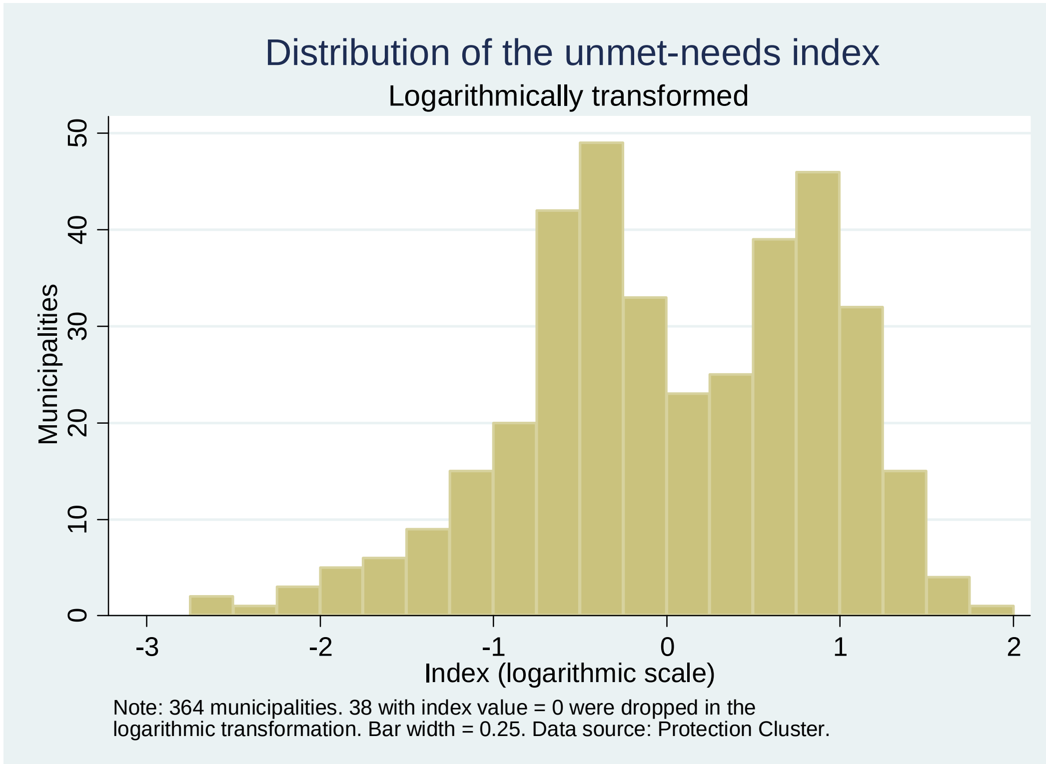
With a mean index value of 4.1 only, the histogram has little information value.



Outliers can be dealt with in different way. A histogram without them can be presented in immediate neighborhood to a list of the individual outliers. Readers may appreciate judging from the sub-index values which component was the major driver in particular cases. In the list of the five neediest municipalities above, it is easy to see that Tacloban City came first because of the high magnitude and intensity that it suffered. No. 5, Capiz, is there because of high values on intensity and pre-existing conditions. Etc.

Alternatively, in the case of a single outlier like Tacloban City, a histogram can be made excluding the outlier, but noting the exclusion prominently with the index value of the outlier.

When speaking to a technically interested audience, it may be fruitful to use a logarithmic transformation of the index. Units with index value = 0 are lost this way, but we would anyway not consider them as genuinely affected. No high-end outliers are excluded. The following histogram is clearly bimodal. Since there is no ex-ante reason for discontinuity, the trough near NeedsIndex = 1 (i.e., log10Index = 0) suggests that different criteria were at work in assessing municipalities with significant needs and those only marginally needy.



Winsorizing extreme values

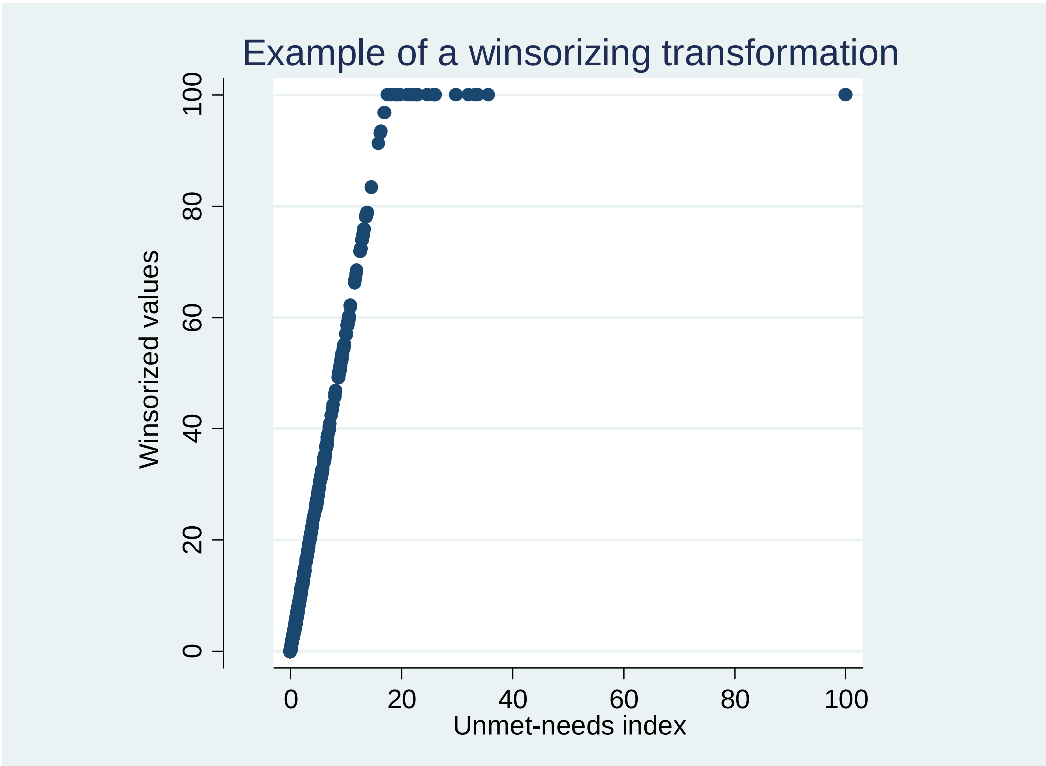
Finally a word on winsorizing, a procedure that re-assigns more in-lying values to units with extreme values. In general, winsorizing is justified only when there are strong reasons to consider those extreme values as understated (at the low end) or exaggerated (at the high end). In this case, the winsorizing should take place in the component that we mistrust the most, before aggregation.

Winsorizing purely for easier visualization is not legitimate and would amount to pseudo-ranking. For example, if we assigned an index value of 40 to Tacloban City, this community would still rank highest on the needs index. But the point is that we believe that its needs are not just greater than those of any other municipality, but, more precisely, they are *three times as great* as those of the second-rank community, Ormoc City (see list above). The index rates communities, rather than merely ranking them.

Technically, winsorizing proceeds by setting all values to be changed to the lowest / highest acceptable inlying value. For the sake of demonstration only, we winsorize the five percent of the highest values. We then re-scale the winsorized values with max = 100.

```
              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------------
UnmetNeedsM~100 float  %9.0g                  Index of unmet needs (multiplicative
                                                aggregation) (re-scaled with max = 100)
NeedsWinsor~100 float  %9.0g                  Index of unmet needs (winsorized, 5% high
                                                values only, re-scaled w. max = 100)
```

This is tantamount to this piece-wise linear transformation:



The mean is now dramatically higher:

```
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
UnmetNee~100 |       402    4.082075    7.818049          0        100
NeedsWin~100 |       402    20.19161    28.48897          0        100
```

We repeat that the truncation of extreme values requires solid reasons and should remain the exception. In general, it is against the spirit of combining ratio-level indicators in a composite measure. Outliers are interesting *because* they are outliers. How far they are separated from the rest of the distribution is one of the things we want to know.

Simulation of the robustness of the needs index

Code

```
************************************************************************************************************
*
* TYPHOON YOLANDA, PHILIPPINES - EXAMPLE OF A QUICK-AND-DIRTY IMPACT MEASURE
*
************************************************************************************************************
* Written by Aldo Benini, for ACAPS. 15 April 2014.
************************************************************************************************************
*
* SIMULATION OF THE ROBUSTNESS OF THE UNMET-NEEDS SCORE TO MEASUREMENT ERROR
*
************************************************************************************************************
* For the conceptual basis, refer to the main body of the note.
*
************************************************************************************************************

************************************************************************************************************
* This code can be used both for summary tables of mean errors and for graphing.
*
* The simulation producing the summary tables, with 200 replications, executed in 38 seconds.
************************************************************************************************************

* VARIABLES USED:
* Three indicators are multiplicatively aggregated:

* 1. The number of affected persons [for the magnitude]
* 2. The proportion of houses destroyed [for the intensity]
* 3. The pre-crisis poverty rate [for the pre-existing conditions]
*
* For simplicity it is assumed that the poverty rate was measured without error.
* For statistical reasons (logit transformation), the housing damage rates were modified as
* 100 percent to 99 percent; zero to one percent.
* For modeling reasons (multiplicative aggregation), the minimum number of affected persons was set to one.
*
* With this modification, the descriptive statistics of the observed values are:

*               storage  display     value
```

```
* variable name   type   format      label      variable label
* -------------------------------------------------------------------------------------
* affectNo0       float  %8.0g                  Affected persons
* totaldamage     float  %9.0g                  Fraction totally damaged houses
* povertyrate     float  %8.0g                  Poverty rate
*
*     Variable |       Obs        Mean    Std. Dev.       Min        Max
* -------------+--------------------------------------------------------
*    affectNo0 |       408    33026.83    56399.22          1     866171
*  totaldamage |       408     .175999    .2650943        .01        .99
*  povertyrate |       408    .3153091    .1267646      .0374     .75461
*****************************************************************************************************

* Some house-keeping:
set more off
timer clear 1
timer on 1

* Set the working directory:
cd C:\[path]


*******************************************************************************************************
* PROGRAM:                                                                                            *
*******************************************************************************************************

* The program part, which the simulate command (below) calls to generate observations with measurement
error.
capture program drop UnmetNeeds /* capture ignores the error if there is no program "UnmetNeeds" to drop */

program UnmetNeeds, rclass
version 12 /* STATA version 12 */

* The program arguments:
args errorMultAffected errorMultDamage
* These are names for the error step-up factors used in the formulas below. The simulate command will pass
values to them.
* Further below, in the simulation part, errorMultAffected will be represented by i, errorMultDamage by j.

* The data file. Adjust path to your computer set-up:
use "C:\[path]\140411_1135_QuickAndDirtyIndex.dta", clear

* Some more house-keeping:
```

```
    * Keep the mean of totaldamage available:
    quietly summ totaldamage
    global meanDam = r(mean)

    * Possibly some redundancy in these "capture drop .." lines (the command is repeated further below).
    * If the variables that are going to be created in the next lines below this block
    * already exist and need to be dropped before the next replication, this does the job.
    * If they don't exist, "capture" prevents the errors from shutting down the program.

          capture drop zAff_*
          capture drop zAffAbs_*
          capture drop AffectedEst_*

          capture drop zDamage_*
          capture drop zDabs_*
          capture drop DamageEst

          capture drop NeedsEst_*
          capture drop NeedsRankEst_*

          capture drop MeanzAffAbs
          capture drop MeanzDabs

          capture drop meanNeedsEst
          capture drop p05NeedsEst
          capture drop p95NeedsEst
          capture drop sdNeedsEst

          capture drop meanNeedsRankEst
          capture drop p05NeedsRankEst
          capture drop p95NeedsRankEst
          capture drop sdNeedsRankEst

* Simulations:
* Create x-number generations of estimated needs for each combination
* of error levels in affected persons and in housing damage levels
* Again note: Poverty rate is assumed to be error-free.


* How many runs?
local noruns = 200
```

```
forvalues k = 1/`noruns' {

* Keep the results exactly reproducible by fixing the seeds of the random variables:
local seedk = 1234 + `k' + 1000 * `errorMultAffected' + 10000 * `errorMultDamage'
* The formula ensures a different seed for each run as long as noruns < 1000.
set seed `seedk'

* Create errors in estimated numbers of affected persons:
* The coefficient 0.3 is the result of trial-and-error, in order to obtain a realistic spread of error
levels.
gen zAff_`k' = rnormal() * (0.3 * `errorMultAffected') /* Error factor */
gen zAffAbs_`k' = abs(exp(zAff_`k') - 1) /* Relative error */

gen AffectedEst_`k' = exp(zAff_`k') * affectNo0


* Create errors in estimated rate of totally damaged houses:
* The coefficient 0.6 is the result of trial-and-error, in order to obtain a realistic spread of error
levels.
gen zDamage_`k' = rnormal() * (`errorMultDamage' * 0.6) /* Error factor */
gen zDabs_`k' = abs(zDamage_`k') /* Absolute value of damage rate logit shift */

gen DamageEst_`k' = invlogit(logit(totaldamage) + zDamage_`k')
* Since total damage is a proportion [0.01, 0.99], the imposition of a measurement error
* must not return a value outside [0, 1]. Therefore, the error has to affect the logit, i.e.
* ln(x / (1 - x)) as an additive term (a multiplicative one would cause sharp reversals
* when zDamage < 0). The modified logit is then inverted back to a proportion.

* Compute the needs scores as the product of affected persons, housing damage rate and poverty rate:

gen NeedsEst_`k' = AffectedEst_`k' * DamageEst_`k' * povertyrate
* Create the ranks
egen NeedsRankEst_`k' = rank(NeedsEst_`k'), field
}

* Compute row-wise statistics:

      * A.: of the errors:
      egen MeanzAffAbs = rowmean(zAffAbs_*)
      quietly summarize MeanzAffAbs
      return scalar MeanAbsEFAffected = r(mean) /* Mean relative error factor in affected persons */

      egen MeanzDabs = rowmean(zDabs_*)
```

```
      quietly summarize MeanzDabs
      return scalar MeanAbsEFDamage = r(mean) /* Mean absolute error in the logit of the housing damage
rate */
      * This value has no intuitive interpretation and will have to be retranslated into a percentage
difference
      * in the final outcome table.

      * B.: of the scores:
      egen meanNeedsEst = rowmean(NeedsEst_*)
      quietly summarize meanNeedsEst
      return scalar MeanMeanNeedsEst = r(mean)

      * For graphing purposes only (5% confidence intervals):
      egen p2p5NeedsEst = rowpctile(NeedsEst_*), p(2.5)
      egen p97p5NeedsEst = rowpctile(NeedsEst_*), p(97.5)

      * This is the statistic of major interest, expressing the amount of uncertainty in response to error
levels:
      egen sdNeedsEst = rowsd(NeedsEst_*)
      quietly summarize sdNeedsEst
      return scalar MeanSDNeedsEst = r(mean) /* Mean standard deviation of the row means of the unmet needs
score */

      * C.: of the ranks:
      egen meanNeedsRankEst = rowmean(NeedsRankEst_*)
      * The mean of these rowmeans always is = ((noruns + 1) / 2). Therefore we do not bother about a
scalar.
      * For graphing purposes only (confidence intervals):
      egen p2p5NeedsRankEst = rowpctile(NeedsRankEst_*), p(2.5)
      egen p97p5NeedsRankEst = rowpctile(NeedsRankEst_*), p(97.5)

      egen sdNeedsRankEst = rowsd(NeedsRankEst_*)
      quietly summarize sdNeedsRankEst
      return scalar MeanSDNeedsRankEst = r(mean) /* Mean standard deviation of the row means of the unmet
needs rank */

capture drop zAff_* zAffAbs_* AffectedEst_* zDamage_* zDabs_* DamageEst_* NeedsEst_* NeedsRankEst_*

end /* End of program set-up */

* Use the program:
* Determine whether for graphing or for error statistics:
* If GraphNotStats == 1, then produces graph; else collects error statistics.
```

```
local GraphNotStats = 1
**********************************************************************************************************
* GRAPHS:
**********************************************************************************************************
if `GraphNotStats' == 1 {

* Set the argument values for the error levels that you want the simulation to use,
* and the results of which you want graphed out. E.g., UnmetNeeds 1 1, or Unmet Needs 3 2, or whatever
integers.
UnmetNeeds 1 1
* These values produce a mean relative error in affected persons of 25 percent, as well as
* a mean absolute error of 8 percent in the proportion of totally damaged houses.
* Such error levels may be unrealistically low for the early stages of assessments.
* To obtain the mean errors at higher steps, first run the simulation tables.

* Graphing only the observed highest-ranked municipalities:

preserve /* Keep a copy of the data table before observations are dropped. Will be restored after graphing.
*/

* How many cases?
local onlyhighest = 20
keep if NeedsQuickRank <= `onlyhighest'
sort NeedsQuickRank

      * Which kind of graph? Ranks = 1; Needs index values = 2
      local graphtype = 2

      * Note: Unable to pass the number of runs as a local. If this is changed, adjust manually in the
twoway code:

      if `graphtype' == 1 {
      twoway (rarea p2p5NeedsRankEst p97p5NeedsRankEst NeedsQuickRank, sort fcolor(gs14)) ///
             (line meanNeedsRankEst NeedsQuickRank, sort lwidth(thick)), ytitle(Simulated rank) ///
             yscale(reverse) ylabel(1 25(25)100) xtitle(Observed rank) xlabel(1 5(5)20) ///
             title(Robustness of unmet-needs index to measurement error) ///
             subtitle(Stability of the ranking in the top twenty affected municipalities) ///
             note("Note: 408 municipalities. Data source: Protection cluster. Quick-and-dirty index
method:" ///
             "Needs = constant * Affected persons * Building destruction rate * Pre-crisis poverty
rate.""Simulation with 200 runs.") ///
             legend(order(2 "Mean" 1 "95% CI" ))
      }
```

```
      else if `graphtype' == 2 {
      twoway (rarea p2p5NeedsEst p97p5NeedsEst NeedsQuickRank, sort fcolor(gs14)) ///
             (line meanNeedsEst NeedsQuickRank, sort lwidth(thick))  ///
             (line NeedsQuickDirty NeedsQuickRank, sort lwidth(medium)), ytitle(Needs index value) ///
             xtitle(Observed rank) xlabel(1 5(5)20) title(Robustness of unmet-needs index to measurement
error) ///
             subtitle(Stability of the index values in the top twenty affected municipalities) ///
             note("Note: 408 municipalities. Data source: Protection cluster. Quick-and-dirty index
method:" ///
             "Needs = constant * Affected persons * Building destruction rate * Pre-crisis poverty rate."
///
             "Simulation with 200 runs. Needs index not re-scaled.") ///
             legend(order(3 "Observed"   2 "Simulated (mean)"  1 "95% CI" ))
      }
      else { /* Do nothing.*/
      }


restore
exit
}

else {
*****************************************************************************************************************
* SIMULATE: Use the simulation command to collect statistics.
*****************************************************************************************************************
      * The simulation itself was already done in the program part above.
      * Therefore only one replication is used here.

      * Creating an empty shell for the collection of simulation results at the bottom of the do file:
      clear
      gen recno = _n
      gen byte AffectedErrorStep = .
      gen byte DamageErrorStep = .
      save CollectSimResults, replace

      * "forvalues" augments the measurement error factors in steps from 0 (no error) to 4 for affected
persons, resp. for housing damage.

      forvalues i = 0/4 {
            forvalues j = 0/4 {
                   simulate   xMeanAbsEFAffected = r(MeanAbsEFAffected) ///
                                    xMeanAbsEFDamage = r(MeanAbsEFDamage) ///
```

```
                                xMeanMeanNeedsEst= r(MeanMeanNeedsEst) ///
                                xMeanSDNeedsEst = r(MeanSDNeedsEst) ///
                                xMeanSDNeedsRankEst = r(MeanSDNeedsRankEst) ///
                  , reps(1) nodots: UnmetNeeds `i' `j' /* Only one replication!*/
            summarize

            tempfile results
                  gen byte AffectedErrorStep = `i'
                  gen byte DamageErrorStep = `j'
                  save "`results'", replace

            use CollectSimResults, clear
            append using "`results'"
            replace recno = _n
            save CollectSimResults, replace

            }
      }
* Compute difference in damage housing proportion, in percent, between the observed mean
* - meanDam from above - and the mean obtained from adding the mean absolute error to the logit.
gen damageMeanDiff = (invlogit(xMeanAbsEFDamage + logit($meanDam)) - $meanDam) * 100

* House-keeping:
label var recno "Record number"
label var AffectedErrorStep "Step factor in errors of affected persons"
label var DamageErrorStep "Step factor in errors of housing damage"
label var xMeanAbsEFAffected "Mean relative error, affected persons"
label var xMeanAbsEFDamage "Mean absolute error in the logit, housing damage"
label var xMeanMeanNeedsEst "Unmet needs score - mean of row means"
label var xMeanSDNeedsEst "Unmet needs score - mean of row standard deviations"
label var xMeanSDNeedsRankEst "Unmet needs rank - mean of row means"
label var damageMeanDiff "Mean percent error, housing damage"

save CollectSimResults, replace

*****************************************************************************************************
* KEY STATISTICAL TABLES                                                                            *
*****************************************************************************************************
* The mean relative error in affected persons. This is the mean of the multiplicative error:
table AffectedErrorStep , c( mean xMeanAbsEFAffected )

* The mean percent error in the proportion of damage houses:
table DamageErrorStep , c( mean damageMeanDiff )
```

```
* The mean unmet-error score, by error step factors in affected persons and damaged housing:
table AffectedErrorStep DamageErrorStep , c( mean xMeanMeanNeedsEst) row col

* Ditto, but the mean of the row standard error (row SD = the SD of the simulations for a given
municipality):
table AffectedErrorStep DamageErrorStep , c( mean xMeanSDNeedsEst ) row col

* Ditto for the ranks:
table AffectedErrorStep DamageErrorStep , c( mean xMeanSDNeedsRankEst ) row col

*End of simulation section.
}

* End of do-file house-keeping:
timer off 1
timer list 1
set more on
```

References

Acaps (2014). Philippines - Typhoon Yolanda. Secondary Data Review [January 2014]. Geneva, Assessment Capacity Project (ACAPS).

Bayudan-Dacuycuy, C. and J. A. Lim (2013). "Chronic and Transient Poverty and Vulnerability to Poverty in the Philippines: Evidence Using a Simple Spells Approach." Social Indicators Research: 1-25.

Betti, G., B. Cheli, et al. (2005). On the construction of fuzzy measures for the analysis of poverty and social exclusion. INTERNATIONAL CONFERENCE TO HONOUR TWO EMINENT SOCIAL SCIENTISTS C GINI AND MO LORENZ. Annals... University of Siena.

Burns, T. R. and T. Dietz (1992). "CULTURAL EVOLUTION: SOCIAL RULE SYSTEMS, SELECTION AND HUMAN AGENCY." International Sociology **7**(3): 259-283.

Campbell, D. T. (1960). "Blind variation and selective retentions in creative thought as in other knowledge processes." Psychological Review **67**(6): 380-400.

IMWG (2014). Extended IMWG Meeting on System-Wide Level 3 Responses [Minutes of the meeting on 20 March 2014, Geneva, Switzerland]. Geneva, Information Management Working Group.

Kienberger, S., T. Blaschke, et al. (2013). "A framework for spatio-temporal scales and concepts from different disciplines: the ‘vulnerability cube’." Natural Hazards **68**(3): 1343-1369.

Kolenikov, S. and G. Angeles. (2004)."The Use of Discrete Data in Principal Component Analysis With Applications to Socio-Economic Indices. CPC/MEASURE Working paper No. WP-04-85." Retrieved 21 January 2005, from https://www.cpc.unc.edu/measure/publications/pdf/wp-04-85.pdf.

Lin, L. I. (1989). "A concordance correlation coefficient to evaluate reproducibility." Biometrics **45**(1): 255-268.

Longest, K. C. and S. Vaisey (2008). "fuzzy: A program for performing qualitative comparative analyses (QCA) in Stata." The Stata Journal **8**(1): 79-104.

NATF. (2014)."2014 Humanitarian Needs Overview: Prioritsation Tool. Technical Note." Geneva, Inter-Agency Standing Committee (IASC) Needs Assessment Task Force (NATF). Retrieved 21 March 2014, from https://assessments.humanitarianresponse.info/guidance#PrioritisationTool.

Pi Alperin, M. N. and P. Van Kerm (2009). mdepriv - Synthetic indicators of multiple deprivation. v2.0 (revised March 2014). Esch/Alzette, Luxembourg, CEPS/INSTEAD.

Royston, P. and S. Kolenikov (2013). shapleyx - Shapley value decomposition [Stata command], http://www.homepages.ucl.ac.uk/~ucakjpr/stata.

Shorrocks, A. F. (2013). "Decomposition procedures for distributional analysis: a unified framework based on the Shapley value." The Journal of Economic Inequality **11**(1): 99-126.
Stata Corporation (2011). Stata Statistical Software, Release 12: Reference Manual. College Station, TX, StataCorp LP.
Thió-Henestrosa, S. and J. Martín-Fernández (2005). "Dealing with compositional data: the freeware CoDaPack." Mathematical Geology **37**(7): 773-793.
Tofallis, C. (2013). "An automatic-democratic approach to weight setting for the new human development index." Journal of Population Economics **26**(4): 1325-1345.
UNOCHA and Maplecroft (2011). 2012 Global Focus Model [27 December 2011]. Geneva, Switzerland and Bath, United Kingdom, United Nations Office for the Coordination of Humanitarian Affairs and Maplecroft.
Verity, A. (2014) "Humanitarian Gap Analysis. Purpose, Definition and Products [Draft, January 2014]." from https://docs.google.com/document/d/1xJuYuv6NXxDklHy9BbxI2WKA40aGR1OpROKQaLDL4VQ/edit?pli=1.
Wikipedia. (2014a)."Absolute deviation." Retrieved 16 April 2014, from http://en.wikipedia.org/wiki/Mean_absolute_deviation.
Wikipedia. (2014b)."Typhoon Ketsana." Retrieved 20 May 2014, from http://en.wikipedia.org/wiki/Ketsana.
Wikipedia. (2014c)."Winsorising." Retrieved 28 March 2014, from http://en.wikipedia.org/wiki/Winsorising.
Yoon, K. P. and C.-L. Hwang (1995). Multiple Attribute Decision Making. An Introduction [Quantitative Applications in the Social Sciences #104]. Thousand Oaks, Sage University Paper.
Zhou, P. and B. Ang (2009). "Comparing MCDA aggregation methods in constructing composite indicators using the Shannon-Spearman measure." Social Indicators Research **94**(1): 83-96.