

Aldo Benini

A note for ACAPS

## Data analysis in needs assessments

### Summary

This note discusses the analysis of data collected during joint rapid assessment, also known as Phase 2 type of assessment, when data is collected at the community level. The analysis plan relies on formats centered on the key issues of affected communities and on their severity as rated by key informants and assessment teams. Some principles, tools and analysis forms are likely suited for other types of assessments as well. The premise, however, is that analysis, like the other Phase-2 assessment activity, happens under severe time pressure. Time calls for rapid condensation and triage from among the theoretically feasible and desirable.

The style of the note alternates between exposition of analytic strategies and their partial demonstration using the data from a recent assessment in Yemen. This data has been reformatted to conform to the data management template introduced earlier. The reader is invited to first familiarize with the document "*A template for managing data in needs assessments*" and its companion Excel workbook. This analysis note too comes with a companion notebook, which readers may use to try out, and improve on, analysis steps.

An effective analysis strategy will balance the "key issue - severity rating" type of data and the local specifics of the assessment with general good practice. The note walks the reader through formal preliminaries into substantive analyses. These are exemplified with the setup needed to compute basic descriptive - of regions, target groups, site categories -, and affected population statistics. We also look into the correlation between priority sectors and issue severity.

For two reasons - time pressure and the attention span of the assessment consumer -, the complexity of issue formulations and severity judgments has to be reduced drastically. The template has built-in machinery to do some of this, by automatically recoding key issues and intervention recommendations with the streamlined sets that the user specifies. In addition, we show how the application of rules (which the analyst *invents* considering the assessment context) further facilitates the importance ordering of key issues. The table on the next page is the outcome of recoding and rule-based treatment of key issues in the Yemen data set.

**Table 1: Summary key issue importance, Yemen**

Sector	Issue	Importance	Sector	Issue	Importance
<b>Education</b>	Education, access	Orange	<b>Protection</b>	Conflict, crime and violence	Orange
	Education, delivery	Yellow		Safety and security	Yellow
<b>Food Security</b>	Food, effective access	Red	<b>Return</b>	Financial	Orange
	Food, quality	Orange		Safety and security	Orange
<b>Health</b>	Health care capacity	Orange	<b>Shelter</b>	Financial	Orange
	Health condition	Red		Safety and security	Orange
	Human resources	Yellow	<b>WASH</b>	Hygiene	Yellow
<b>Livelihood</b>	Financial	Orange		Sanitation	Yellow
	Human resources	Yellow		Water	Red

*Legend: Red = high importance; orange = medium importance; yellow = low importance.*

Note: Condensed from 59 issue categories with 984 key informant severity ratings from 43 sites.

Finally, we give stopping rules for when the analysis should end. The sooner it ends, and the more parsimonious its output is, the more time will remain for the assessment team to discuss tentative findings among themselves and with stakeholders. This needs to be balanced with appropriate complexity and with transparency, preferably through a technical appendix, of how results were obtained.

A number of practices are suggested by the experience made hitherto:

- Although the data structure calls for a relational database, management and analysis within spreadsheets is feasible. The Pivot table is the workhorse accomplishing most of the tasks.
- Severity ratings produce data of a kind that prohibits simple averages. The frequently seen charts based on the arithmetic mean for sectors and groups are misleading. Valid substitutes are available.
- The analysis is strengthened when not only issues and their severity, but also priority sectors are recorded from key informants.
- The analysis should be driven by tables, much less by visuals. Chart production delays analysis; few charts are needed, and most of these only at the end. Maps are the exception; they are helpful from early on in the process.
- Produce many tables; inspect them closely; discard most; interpret, document and edit the few retained!
- Recoding - not a household word in the humanitarian information community - is key to complexity reduction and rapid production of hypotheses and tentative findings, and finally to output that the consumer can handle cognitively.

We recognize that this analysis plan is largely untested. It may also become irrelevant if major premises of the Phase-2 type of assessment are changed:

- First, the punishing two-week deadline may be taken more liberally. This would reduce pressures, permitting longer and deeper exploration as well as richer reporting.

- Second, the format and the ensuing data architecture may be recast in ways different from a key issue-based approach. In the final section of this note, we hint at other potential approaches.

Some of these differ in small degree only from the practice in Yemen, such as by having the assessment teams rate the severity of all the issues in a sector. A more radical approach would separate the format for the priorities that the community expresses from the one that records the team's own evaluations, reuniting them only in a subsequent step. Firm recommendations are not feasible at this point; more debate and experience need to inform them. What was done in Yemen is not an optimal, but a satisfying approach.

Regardless of the changes that future assessment formats will undergo, we believe that some of the tools and devices demonstrated here may be of use even in vastly different circumstances.

# Contents

Summary .....	1
Acknowledgement .....	6
Introduction.....	7
[Sidebar:] The limits of severity ratings .....	8
Demonstration workbook.....	11
The data source: Yemen.....	11
Structure .....	11
Variability .....	12
Using the workbook .....	12
Analysis program .....	13
Preliminaries .....	13
Data cube .....	13
Use of the substantive tagging variable .....	13
Pivot table templates .....	14
Substantive analyses .....	15
Basic descriptive tables.....	15
Affected populations.....	18
Priority sectors, key issues and severity ratings.....	20
[Sidebar:] The measurement of priorities, issues and severity .....	20
Visual inspection.....	25
Recoding .....	30
[Sidebar:] How much information is lost in recoding?.....	38
When does analysis end? .....	40
Outlook .....	41
References.....	43

## Figures and tables

Figure 1: A substantive use of the include-record tag .....	14
Figure 2: Key issue severity ratings, Yemen, by sector.....	24
Table 1: Summary key issue importance, Yemen .....	2
Table 2: Data at three levels: 43 sites, 344 site-sector combinations, 1,087 problems ....	14
Table 3: 272 severity-rated site-sector combinations, 1073 problems, in 8 sectors .....	15
Table 4: Sites by region and target groups.....	16
Table 5: Communities with priority needs, by target group .....	17
Table 6: Population statistics of affected communities, by target group.....	19
Table 7: Pivot table drill-down example.....	19
Table 8: Key issues recorded, by sector and sector priority .....	23
Table 9: Pivot table of key issues, by sector, filtered to priority sectors and highest severity score .....	26
Table 10: Pivot table of key issues, enlarged to severity score 3 and 4 (segment).....	27
Table 11: Gradual expansion of key issue inspection.....	28
Table 12: Severity ratings of key issues vs. those of synthesis records.....	29
Table 13: Agreement between severity ratings and priority judgments regarding sectors	30
Table 14: Example of a recoding scheme - key issues in the shelter sector .....	32
Table 15: Key issues, recoded, by sector and severity .....	33
Table 16: Same table, with problems restricted to priority sectors.....	34
Table 17: Key issues, recoded and color-marked for attention .....	35
Table 18: Documenting how the recoding was done (segment).....	37
Table 19: Cross-tabulating recoded key issues and recommendation - food security .....	38
Table 20: Calculating the entropy of a key issue distribution .....	39

## **Acknowledgement**

The database of the "Joint Rapid Assessment of the Northern Governorates of Yemen" (ACAPS 2011) was created by a team led by Wolfgang Gressmann. For demonstration purposes, I reformatted the main data table. Any errors, factual or of interpretation, that may have occurred in the process are mine.

Aldo Benini

## Introduction

This note is for teams working in rapid needs assessments. It discusses the analysis of data collected during Phase-2 of an assessment. This phase is variously known as "multi-cluster rapid needs assessment" (NATF 2011b) or as "Phase 2 - Sustaining lives" (Garfield, Blake et al. 2011). Some of the analysis concerns matter also for other phases or for other types of needs assessments that do not follow the four-phase pattern. The analysis plan proposed here is best suited for assessment formats centered on key issues and their severity ratings, as developed by the initial and rapid assessment toolbox (IRA) (IASC 2009a, 2009b), the Joint Education Needs Assessment toolkit (Global Education Cluster 2010: Tool 8) and used by ACAPS in Bolivia and Yemen in 2011.

### Phase 2 of rapid assessments

Phase-2 data is collected<sup>1</sup> in sites or communities selected under a purposive-sampling plan. Such samples generally are small. This is illustrated by several recent assessments. The "Joint Rapid Assessment of the Northern Governorates of Yemen" visited 43 sites. An assessment in Bolivia covered 18 municipalities affected by four different kinds of disasters. In Bangladesh, 63 groups displaced or marooned by flooding were visited. The assessment of over 300 local government areas in the flooded Punjab Province in Pakistan in 2010 belongs to the rarer larger kind.

The information available in Phase 2 generally is more detailed than in Phase 1. This phase is known as "Preliminary scenario definition" or "Phase 1 - Saving lives". Its assessment efforts focus on coarse impact estimates of affected populations by region, social group and sector. Phase 2 descends to affected communities or even to distinct groups within communities. It elicits from them a variety of information formulated as priorities, preferences and/or problems. At the same time, the sector focus is maintained, notably by the way problems are elicited and recorded. By contrast, region and target group are adjunct entities, absorbed from context and specific questions.

Despite greater detail, small purposive samples still pose strong limitations. Those on precision, population-weighting and generalization are foremost. Rates and differences by any categorical variable will be sensitive to measurement error and categorization behavior. The analysis should therefore be sparing, avoiding breakdowns into ever smaller subgroups, and stressing the limitations of the numeric findings. Diversity has as much analytic value as modes and means do.

### One of many possible architectures

In theory, many data architectures are feasible. Practically, traditions form and solidify, inspired by important policy decisions, achievements, or lack of alternatives. This is well illustrated by the assessments that ACAPS led in Bolivia, Bangladesh and Yemen. These assessments had in common a data generation style centered on "key issues" and the severity of their corresponding needs. Key issue severity rating goes back to the Initial Rapid Assessment (IRA) toolbox developed, between 2007 and 2009, with approval from

---

<sup>1</sup> We follow in this note and elsewhere the consistent abuse of the term "data" as a collective singular and therefore use expressions like "the data is ..", "all this data ..", despite the obvious Latin plural form.

the IASC. It is therefore likely to strongly influence the design also of future assessments. All the same, we emphasize that other assessment approaches, necessitating different data architectures and analysis plans are feasible. For example, Phase-2 assessments could be built around a set of humanitarian indicators, in the form of broadly estimated rates of destruction, displacement and loss of essential services.

### **Problems with "problems"**

Apart from high-level endorsements, "key issues" have a substantive rationale. Sectors by themselves are too broad as concepts for response planners to meaningfully relate to assessments. It is specific issues and needs that help situate response options.

However, the way severity-rated "key issues" (or "problems", as they are also called) are analyzed itself can be problematic. This note, therefore, needs to speak also to the "problems of problems" and the correct ways of handling severity scores. The major risks for producing invalid or low-value findings are:

- The IRA-inspired format (IASC 2009a) has spawned tables with hundreds of records and ratings; regardless, the number of truly independent observations is as small as the number of assessed sites. Longer lists of problems, therefore, do not make the findings more robust. They can, however, support more in-depth interpretations.
- The "key issue"-centered format is tempting for the analyst. It invites all manner of cross-tabulations in which the arithmetic mean of the severity ratings is displayed as the key statistic. Unfortunately, this operation is illegal; ordinal ratings have no intelligible mean (see sidebar below).
- Similarly, problems and recommendations are many; thus they feed complex lists and charts. These may be helpful for local purposes, in looking up what the assessment had to say about site X and its sector Y problems. In order to effectively communicate a Phase-2 summary to the national and international audiences, this complexity has to be reduced.

To put a figure on this challenge: The key informants in 43 sites in Yemen had the teams note 1,033 problems, which were recorded in 59 problem categories. They were mirrored in 63 recommendation types. Together we find 146 distinct problem-recommendation pairs. This analysis note emphasizes features that deal with this complexity.

---

### **[Sidebar:] The limits of severity ratings**

The key issues and their severity ratings are at the center of recent assessment formats. They are attractive. The color scheme is intuitively understood by interviewing teams and by key informants who look over their shoulders. Formulas translate checkmarks in numeric scores. The scores are easy to average by issue, sector, target group or region.

Unfortunately, the statistics and the comparisons built on them are not valid. This sidebar corrects this error early on in this note.

Severity ratings produce ordinal data, of the kind "White is brighter than gray; gray is brighter than black. There are many white things, many gray things, and many black things. All white things are

brighter than any gray or black things. But we cannot say by how much." Etc. for gray things vs. black things.

The mean of the scores of an ordinal variable does not have a meaningful interpretation. Working with the mean of severity scores is therefore misleading.

A simple example may demonstrate this. Let us assume that the assessments of two affected communities returned two issues that were identically categorized following these key informant statements:

Community A:

*"Pests have destroyed part of the rice harvest. Moreover, ever since the local textile mill was shut down, many of the landless families here have lacked the money to buy food."*

Community B:

*"The irrigation system was damaged; our farms produced little this year. Many people have found work in towns; they are sending food and money to their families."*

These verbose accounts were coded and scored on the usual color scale, from green ("relatively normal"; numeric value = 1) to red ("severe"; 4) (NATF 2011a: 24, for the complete definitions):

Problems identified	Severity of needs				Score
<b>Community A</b>					
Level of local food production (food availability)			x		3
Purchasing power (food accessibility)				x	4
<b>Community B</b>					
Level of local food production (food availability)			x		3
Purchasing power (food accessibility)	x				1

If we average the scores by issue, food availability comes out as 3, food accessibility as 2.5. If the observer then abstracts from local context, the mean scores suggest that food availability is the greater problem. But this interpretation may disguise a keen local entitlement crisis.

This example simplifies to the extreme. However, the larger the number of sites over which the severity scores of an issue are averaged, the less our ability to consider local context. The mean misleads.

The median is, in theory, a legitimate alternative. However, as demonstrated further below, there is a "problem with problems". Below a certain recognition threshold, problems are not communicated, and are not recorded. This accounts for the rarity of "green" entries. The scarcity of recorded minor problems artificially drives up severity score medians - producing grade inflation.

Some of the remainder of this note is about what can be done in this predicament.

While we continue to work with this issue and severity rating-driven set-up, we again point out that there are other approaches to needs measurement, with different analysis challenges. An analysis plan for assessments using a combination of adapted HESPER type of questions (WHO and King's College London 2011) and humanitarian indicators was outlined in an earlier report (Benini 2011a: 27, 34-38). The final section of this note hints at yet other possibilities.

### **What is analysis?**

By and large, analysis in this context is quite mundane. It means producing two- or three-way cross-tables, using the Excel workhorse known as the Pivot table. Since an almost limitless number of such tables are conceivable, an analysis plan should privilege a small repertory of tables considered basic. Other types will be called for by the unique context and evolution of the disaster. More will be produced when the analyst detects something worth investigating in greater detail or in different arrangement.

Less trivially, we strongly emphasize "recoding" in this note. Recoding is not a household word in the assessment community, but it is an important tool for managing complex categories as well as for sharpening the analysis. Recoding facilities are pre-formatted into the data entry template.

These remarks give away that we prefer an analysis style driven by tables. We should make a lot of tables, and in Excel we can do so fast. We should discard most after the first inspection. The few that we retain for further use should be interpreted carefully. We are in the realm of descriptive statistics; here charts seldom enlighten any better than a close reading of the underlying tables does. Chart production slows down the process; most can wait until we straighten out table interpretation. Raw maps do let us detect patterns if GIS assistance is on hand; very few are later needed in print quality. The needs of the analyst are not the needs of the consumer.

What the consumer expects of the analyst is relevance, novelty (in the sense of non-trivial communications) and presentation reduced to a level of complexity that fits in his (generally very short) attention span.

### **What to expect**

The rest of this note proceeds as follows: The analysis program and the set-up of selected Pivot tables are demonstrated using the Yemen assessment data. An Excel workbook is provided; the next section explains its structure. The Yemen data has been recast in the format of the generic data management template developed for Phase-2 data; we assume that the reader is familiar with the companion note and workbook on data management.

The next chapter develops the analysis program. Its preliminaries elaborate on the use Pivot tables and tagging variables. Next, substantive analyses are exemplified for needed basic descriptives, affected populations, as well as the interplay of priority sectors and key issue severity ratings, with deliberate preference for close inspection of tables over charts. A lengthy section is devoted to recoding problems and recommendations, an operation that we believe is key to successful analysis, particularly under time pressure. In the same chapter, we demonstrate how the analyst, by inventing and applying clear and transparent rules of interpretation, produces compact findings that the assessment consumer can absorb. We give stopping rules for when the analysis should end. In the final outlook section, we re-emphasize the needs for parsimony and triage under the tough condition of Phase-2 work .

## Demonstration workbook

### ***The data source: Yemen***

The data of the above-mentioned Yemen assessment was made public. We use it to demonstrate recommended analysis forms. The data has, with few exceptions, not been altered, but we have rearranged it. It now conforms to the generic data entry template discussed in the previous note. The reader looking for explanations of the table structure is referred to that document.

The original data table had 1,380 records, whereof 1,087 were key issue instances, and the rest were based on summary evaluations (called "syntheses") of the particular site-sector combinations under which one or several issue instances had been recorded.

### ***Structure***

The resulting data table has records at three different entity levels - the affected community (the site), the site-sector combination, and the key issue. Issues are called "Problems" in the database.

As described in the data management note, for each site 65 records are reserved. The top row of this block is a site level record. For each of the eight sectors for a given site, seven rows are reserved for problems. An additional row holds the synthesis record. The table has 2,795 records, including the non-informative synthesis and problem records. Other important features of the data table include:

- The record numbers (column 1) have been made static; the formulas that created them automatically were replaced with their values. The table now is sortable.
- Tagging variables make it easier to navigate and filter to subsets of records that are of interest during the analysis. Separate tags are provided for the *site* and *synthesis*-level records. The "*Keep\_record*" tag says "Yes" if the record: 1. is a site-level record, or 2. a "Synthesis" record, or 3. an informative problem record. We tag "Synthesis" records to be kept even if no problems were recorded for this site-sector combination. This keeps these site-sector combinations available for certain types of analyses. In the rightmost column, another tagging variable has been provided, this one for a substantive purpose, which we explain below.
- The data table, as reconstructed, includes the recoding facilities as provided by the data entry template. We also provide a second table from which empty problem-level records were deleted and all formulas were replaced with their results. If you wish to export the data to some other application, we recommend using this sheet, named "Data\_Static". This table has 1,474 records.

The sheet "Variables" lists the variable names and labels, together with comments. Many overlap with the default variables of the template. Of special note, key informants at every site designated three sectors as priority sectors. Severity ratings were given to problems and to sectors (under the syntheses).

The categories used in the "Problem" and "Recommendation" variables resulted from a standardization process that took place during data entry. The database holds no transcriptions or comment fields recording the original verbose key issue formulations.

### ***Variability***

Some of the variability in the Yemen data will be used to illustrate analysis considerations further below. The assessment teams returned data on 43 sites, with an unknown number of sites in one district bunched into a combined site. Four types of affected people ("target groups") are distinguished. Problems were elicited in eight sectors; when noting priority sectors, the category "cross-cutting" was also used, if rarely.

As mentioned above, the categorization of problems created 59 types while there are 63 types of recommendations. These figures include syntactic differences, such as when "child labor" is spelt differently. There are 146 distinct problem-recommendation pairs. This diversity is considerably reduced when we look only at sectors that are priorities for the assessed communities. This subset of 420 problem instances uses only 74 such pairs.

### ***Using the workbook***

Users intending to practice with the data may want to save the workbook in a different name.

The workbook has 18 sheets and a table of contents with hyperlinks to the sheets. An easy approach is by first looking at the sheets "Variables" and "Pivot\_DataOverview".

As mentioned, there are two versions of the data table - one with active formulas supporting the recoding effort, the other static and meant for exporting data to other applications and sharing with outsiders. We assume that you will be working with the active version. Its data range has been named "Database", for convenience in Pivot tables. The definitions of other named ranges are listed in an auxiliary sheet if you wish to look them up; normally you will not need it.

As explained in the data management note, the problem and recommendation categories used are listed in separate sheets for each sector. Changing these categories at this point has no effect. However, we encourage users to experiment with recoded categories, in column 3 and 8 of the sector-wise sheets, and to see how the changes are reflected in Pivot tables that use the recoded problems and recommendations.

Most of the demonstrations use site, sector, key issue and severity rating variables. The workbook holds only one sample Pivot table linking problems and recommendations (the sheet "RecommendExample"). The formalities of analyzing recommendations are nothing special. The challenge is in recoding them in ways that simplify enough, yet maintain non-trivial substance.

# Analysis program

## *Preliminaries*

### **Data cube**

We continue to work with the concept of a flexible data cube, as recommended for Phase-1 and 2 data management (Benini 2011b). In Excel, the "cube" concept is implemented through Pivot tables. Two more features are useful:

- A tagging variable which allows us to condition the selection of records in response to substantive information. We find it in the same place as in the template - in the rightmost column of the data table.
- The drill-down feature of Pivot tables themselves. Double-clicking on a data cell in the Pivot table throws up, in a new sheet, the segment of the database by values of the row and column variables that define that cell.

With one exception, the tables that we demonstrate are not edited for reports. They are raw Pivot tables, with the filter settings shown. For clarity, we edited column headers in some tables. Editing tables to presentation quality, of course, takes considerable work, but this is not a concern of this note.

### **Use of the substantive tagging variable**

In Yemen, key informants were asked to designate the three sectors in which the needs of their communities had the highest priorities. We create a tag for those problem records as well as synthesis records whose associated sector was a priority for the community. This excludes records from non-priority sectors as well as the site-level records. A formula like

$$=IF(OR(RC22=RC16, RC22=RC18, RC22=RC20),1,0)$$

translates this logic. If any of the three priority sectors (column 16, 18, 20; these are site-level attributes) match the sector (column 22; a synthesis or problem-level attribute), we tag the record. We note the interpretation by changing the field header to "IsPrioritySector". The following screenshot shows an instance of the formula and the change of results between a non-priority sector (shelter for this community) and a priority one (food security).

Figure 1: A substantive use of the include-record tag

R26C29		=IF(OR(RC22=RC16, RC22=RC18, RC22=RC20),1,0)	
22		29	
1	Sectors and sect	Include record in Pivot table	
2	Sector	IsPrioritySector	
26	Shelter	=IF(OR(RC22=RC16, RC22=RC18, RC22=RC20),1,0)	
27	Shelter	0	
28	Food Security	1	
29	Food Security	1	

We will use this tag in Pivot tables where excluding non-priority sector information is a needed step. If this tag is needed for a different purpose, the formulas and the field name will have to be modified accordingly. It is almost needless to say that the inclusion tag can be used also in the data table itself, in sorting or filtering.

### Pivot table templates

Using the formal and substantive tags, we build two basic Pivot tables in a sheet that we name "Pivot\_DataOverview". These tables give a succinct overview of the data situation, as the captions summarize:

Table 2: Data at three levels: 43 sites, 344 site-sector combinations, 1,087 problems

Keep_record	Yes
IsPrioritySector	(All)

Count of RecordNo	Synthesis_record		Grand Total
	Site_level_record No	Yes	
No	1087	344	1431
Yes	43		43
<b>Grand Total</b>	<b>1130</b>	<b>344</b>	<b>1474</b>

**Table 3: 272 severity-rated site-sector combinations, 1073 problems, in 8 sectors**

Keep_record	Yes
Site_level_record	No
IsPrioritySector	(All)
Severity	(Multiple Items)

Sector	Synthesis_record		Grand Total
	No	Yes	
Education	160	38	198
Food Security	108	36	144
Health	140	37	177
Livelihood	136	36	172
Protection	133	35	168
Return	68	17	85
Shelter	140	35	175
WASH	188	38	226
<b>Grand Total</b>	<b>1073</b>	<b>272</b>	<b>1345</b>

**Note:** The "multiple items" in the Severity filter appears because blanks and N/A were excluded.

Apart from their overview function, these tables offer convenience templates. Copies put in other sheets serve as starting points for almost any other Pivot table. These and other fields can then be moved or dragged from the Pivot table wizard as needed.

### ***Substantive analyses***

We make a distinction here between basic descriptions of the key assessment dimensions - region, groups of affected persons, needs sectors - and analyses that include the problem, severity rating, and recommendation variables. Also, as the analysis progresses, recoding the diversity of problems and recommendations into smaller category sets adds new challenges.

### **Basic descriptive tables**

As a general principle, it may be productive to scout for relationships between two kinds of distributions. The first kind is of combinations of pre-disaster traits, notably region and social group (the latter can be the result of the disaster itself, at least in part).

The second is about the needs occasioned or deformed by the disaster, or about those needs combined with some other disaster outcome, such as origin, host and camp communities.

## Two illustrations

Here are two sample tables. The first is a simple cross-table using site-level records. Note that, apart from the site level and the usual "Keep\_record", no other tags have been specified. The table could of course be enhanced with population figures, but this parsimonious version lets us see a basic structure at a glance.

**Table 4: Sites by region and target groups**

Keep_record	Yes					
Site_level_record	Yes					
IsPrioritySector	(All)					
Severity	(All)					
Target_Group						
Count of RecordNo	up					
Governorate		Host Communities	Other conflict-affected persons	Returning IDPs	Vulnerable IDPs	Grand Total
Al Jawf		2	4		2	8
Amran		5			8	13
Hajjah		5			4	9
Sa_ada				6	3	9
Sana_a					4	4
<b>Grand Total</b>		<b>12</b>	<b>4</b>	<b>6</b>	<b>21</b>	<b>43</b>

The table produces basic findings and puts important questions into a reminder list for the further analysis. Host communities were assessed only in three of the five governorates - are there none in the others? The types of "other conflict-affected" groups, recorded in one governorate only, will need to be explained. Etc.

A map will be helpful for the analysis and more powerful in the report than a barren table.

The second table combines elements from two Pivot tables:

- The first part uses substantive information on the sectors. It is pulled from the synthesis (= site-sector) records. The priority sector tag is active and filters to site-sector instances that the affected communities considered priorities.
- Since the row-wise totals of site-sector instances have no meaningful interpretation, they are suppressed. Instead, the cell frequencies can be compared to the total number of assessed sites for each target group. This information is simply a transposed, static copy of the bottom row of the previous table.

**Table 5: Communities with priority needs, by target group**

Keep_record	Yes	<input type="button" value="v"/>					
Site_level_record	No	<input type="button" value="v"/>					
Synthesis	Yes	<input type="button" value="v"/>					
IsPrioritySector	1	<input type="button" value="v"/>					
Severity	(All)	<input type="button" value="v"/>					
<hr/>							
Count of RecordNo	Sector	<input type="button" value="v"/>					
	Education	Food Security	Health	Livelihood	WASH	All assessed sites	
Target_Group	<input type="button" value="v"/>						
Host Communities		10	5	5	9	12	
Other conflict-affected persons	1	2	3	1	4	4	
Returning IDPs	3	5	1	3	6	6	
Vulnerable IDPs	2	17	14	8	11	21	
<b>Grand Total</b>	<b>6</b>	<b>34</b>	<b>23</b>	<b>17</b>	<b>30</b>	<b>43</b>	

This combined table thus lets us compare frequencies of sector priorities to the number of sites with particular target groups.

### No percentages

Note that we do not use percentages in this and other tables. The sample is small, and some categories are represented in tiny numbers. Education is a priority sector for six out of the assessed 43 communities. One of six returning IDP communities mentioned health in these terms. A change of one in the numerator would drastically affect the proportion.

The table, therefore, has to be inspected visually, cell by cell. As a discovery tool, conditional formatting or heat maps are legitimate, but with small samples like this they cannot signal robust findings. Where proportions differ considerably, robust patterns may be conjectured. Most drastically, three of the eight sectors - protection, return, shelter - were prioritized by none of the Yemen communities. There must be reasons for this absence. Among the sectors that were priorities for some, health-related needs appear to be more acute for vulnerable IDP communities than for their hosts. Intuitively this makes sense, but one wishes that someone familiar with the context explain the whys and hows.

### Cycling through two-way tables

Again we stress that in different contexts different tables will best serve this part of the analysis - as "cube"-inspired descriptions of how the three basic dimensions interact. As a general principle, it can be helpful to cycle through the sequence of two-way tables, as in

*(region X target group), (target group X priority sectors), (priority sectors X region),*

or even four:

*(region X target group), (target group X site category), (site category X priority sectors), (priority sectors X region).*

This is a disciplined means to build, confirm or modify hypotheses from which to start more detailed pattern investigations.

### **Are these relationships robust?**

Further steps, notably in preparation for regional chapters and appendices in the report, may call for three- and four-way tables. These meet lookup functions and finer-grain descriptive needs of the analyst intimately familiar with the data structure. For assessment consumers with limited absorption capacity, they may not be suitable. To them, patterns revealed in two-way tables will often form the major findings that get across.

Statistical programs offer routines that facilitate the discovery of patterns across several categorical variables. Cluster and multiple correspondence analyses are favorites among researchers bridging qualitative and quantitative approaches. But the small-sample limitations do not go away. These results remain highly sensitive to sample changes, missing data and measurement error. Excel does not offer these features, but it lets us do careful work with its fast and flexible tables.

### **Affected populations**

Descriptive statistics using population figures may come in almost infinite variety, depending on particular questions and interests. In any event, a plausibility check on the claimed figures may be in place. Human settlement research has found a lognormal distribution of population size in many regions. Yet, disaster effects readily upset the pre-existing pattern. For example, we may find that only some larger towns are hosts to IDPs; or that IDPs disperse into many small groups, but also congregate in a few large camps. The spread of settlement size by region or target group will therefore be informative, both for its central tendencies and its discontinuities.

### **From summary table to detailed list**

This example uses the Total Population variable several times, for sums, minima, means and maxima. Excel does not offer medians or other percentiles in Pivot tables<sup>2</sup>. Again, note which tags are active, and which are not.

---

<sup>2</sup> Excel users can calculate tables with medians and other descriptive statistics, although with fewer factors and filters than Pivot tables allow, using *SSC-Stat*, a free Excel add-in offered by Reading University, at [http://www.reading.ac.uk/ssc/n/software/sscstat/helpfile/ht\\_start.htm](http://www.reading.ac.uk/ssc/n/software/sscstat/helpfile/ht_start.htm). This tool offers a host of other applications as well, in data manipulation, visualization and analysis. Highly recommended.

**Table 6: Population statistics of affected communities, by target group**

Keep_record	Yes					
Site_level_record	Yes					
Synthesis	(All)					
IsPrioritySector	(All)					
Severity	(All)					
<b>Values</b>						
		Count of RecordNo	Sum of Total_population	Min of Total_population2	Average of Total_population2	Max of Total_population2
<b>Target_Group</b>						
Host Communities		12	167,500	2,000	13,958	70,000
Other conflict-affected persons		4	31,760	470	7,940	14,000
Returning IDPs		6	11,830	315	1,972	4,200
Vulnerable IDPs		21	410,196	350	19,533	184,509
<b>Grand Total</b>		<b>43</b>	<b>621,286</b>	<b>315</b>	<b>14,449</b>	<b>184,509</b>

This leads to new questions. The largest IDP site is almost ten times the group mean. We use the drill-down to inspect the distribution. This is a segment sorted on population size:

**Table 7: Pivot table drill-down example**

Site_Name	Site_category	Urban_Ru	Total_populat	Total_IDPs	Gender
Ta_ashar alirak	Community in the open	Rural	350	350	Female
Ta_ashar alirak	Community in the open	Rural	350	350	Male
Al Batn	Village_part of town	Rural	400	90	Male
Al-baten Al saeed	Village_part of town	Rural	400	90	Female
Shatea_albard	Village_part of town	Rural	462	112	Male
Sumin	Village_part of town	Rural	2,000	135	Male
Almoslahakat	Village_part of town	Rural	2,800	400	Female
Bir Aidh	Village_part of town	Rural	3,000	500	Male
Al-wazeer	Village_part of town	Rural	3,500	1,200	Male
Bait Zood	Village_part of town	Rural	3,800	350	Male
Al Masrak III	Camp	Rural	4,000	4,000	Female
Al Hadaba Sharkiah	Village_part of town	Urban	5,000	400	Male
Al Masrak I	Camp	Rural	9,000	9,000	Female
Dhabr Kant	Village_part of town	Rural	10,955	300	Female
Alhboba-BeirHirab	Village_part of town	Urban	11,680	3,240	Male
Different sites	Village_part of town	Urban	12,300	5,130	Female
Dares	Village_part of town	Urban	24,345	1,690	Male
Dares	Village_part of town	Urban	24,345	1,690	Female
Kaa Alswadin	Village_part of town	Rural	42,000	1,400	Female
Raydah	Village_part of town	Rural	65,000	2,100	Female
Beni Al Harith	Village_part of town	Urban	184,509	5,130	Male

Beni Al Harith is not the largest IDP site, but the largest host community and so much larger than second-largest host that it may be a class of its own. This is a substantive question that cannot be decided on analytic merits. We also find two sites entered twice, by gender. This calls for clarification - what do the figures mean? - and also questions the

data model. Do we avoid double counts by tagging to one gender? If so, which? And finally: How was the line drawn between IDP and host community assessments?

### **Questions from table inspection**

We mention these questions for general reasons only, to show how just a few table operations send surprises and new challenges bursting into the analysis train. They should be met, or at least noted, before we advance detailed results building on the variables in point. For example, sector-wise totals will be meaningless unless we know in which cases to use the total population figures, the IDP figures or both together. If straightforward solutions are not available, it may be better to exclude outliers and special cases from the statistics in point, and devote short descriptions in sidebars to each of them.

The larger question is whether any and, if so, which assessment statistics should be population-weighted. On first thought, the answer is no, because of the purposive nature of the sample as well as the extreme range of site populations. Direct population-weighting would obliterate the influence of the smaller communities. On second thought, total disregard for population size is not satisfactory either. There are reasons why some sites are big, some are medium, and some are small or very small. These reasons are likely correlated with the needs of the affected populations. The size distribution may reflect assessment design or on-spot decisions, such as when there is uncertainty as to how to treat sampled villages and their outlying hamlets, or neighborhoods within cities.

Prudence in this situation advises separate outlier presentation, as suggested, or classification by population size or IDP figures, in the best judgment. Size classes will have to be few, with bounds at major discontinuities. This is not population weighting, but rather stratification on population size. The most important information to be broken down by these strata concerns the priority sectors.

### **Priority sectors, key issues and severity ratings**

The core of the assessment analysis, as practiced in several recent country assessments, gravitates around key issues - interchangeably called "problems" - and the corresponding severity ratings. In addition, some assessments - Yemen, in modified manner Bangladesh - let communities determine priority sectors. The logic and interaction of these variables are therefore critical to valid analysis. This is particularly true of key issues and severity ratings, given their canonical status in IASC documents. Some of the analyses made of severity ratings in Bolivia and Yemen violated statistical assumptions as well as common sense. In this sidebar, we talk about some more abstract methodological aspects.

---

### **[Sidebar:] The measurement of priorities, issues and severity**

This sidebar is fairly technical and lengthy. Readers not interested in this type of conceptual inquiry may skip it without loss of understanding of subsequent sections.

#### **Priority sectors**

When communities are asked to determine which of the sectors have priority for them, a number of challenges may arise. The sector categories of the humanitarian community will in part be

foreign to local culture and language; the mapping of their expressed needs to sectors may be done by the assessment teams, on extraneous considerations. Those and other measurement challenges, however, need to be addressed *before* the data analysis. Here, we are concerned with analysis - with the fact that the priority response produces data on genuinely ranked items. There is only one first-priority sector.

Practically, not all sectors are ranked. Three appears to be common practice. It is also the number of priority concerns elicited in the Hesper Scale instrument (WHO and King's College London 2011). The rest of the sectors remain unranked; all that can be said about them is that they are not among the first three priorities. It is helpful to score by inverse rank, with "3" for the first priority sector, etc. and zero for all unranked ones.

This data is ordinal. Statistically, the mean of the rank score has no interpretation. However, the scores (as inverse ranks) may be summed in the understanding that communities "vote" for multiple candidates to be their priority sectors. These votes are weighted in the above manner. This system, known as "Borda counts" (Borda 1781; Benini 2011b; Wikipedia 2011a), is considered, if not an optimal, then a satisfactory ranking system, particularly on account of its simplicity.

If, in this specific situation, summing the ordinal scores is legal, then, for equal numbers of voters over all items being considered, trivially also the mean score is legal. The key point is that for every community assessed, each of the non-zero scores is given to one sector only. Sometimes one finds priority rankings called "cross-cutting" [= across sectors]; that and similarly vague terms undermine the validity of Borda counts. Items presented to only part of the voters - e.g. when only IDPs are asked about the priority of "return" - too can distort results because the counts for the sectors are computed on different subsets of the purposive site sample.

The Borda count property makes it important to have ranked sector priorities besides the key issue and severity ratings. Note that the variables thus created - either three multinomial ones, "first priority sector", second, third - or eight ordinal ones, "Food security - priority score", shelter priority score, etc. - are properties of the assessed sites. In other words, they are not properties of the sectors, nor of the site-sector combinations. The latter may inherit the information; practically, this is done by copy-paste in the data table.

### **Key issues**

"Key issues" present multiple-response data. Sites can express none, one or several issues under every sector to which the interviews are guided. Up to seven per sector could be recorded in the questionnaires used in Bolivia and Yemen.

However, the similarity with multiple response data is only in form, not in generation. There is no pre-defined problem or key issue list, just as there is no agreed-upon subsector list. The structure of the data generation process is poorly understood. There is a small literature on agenda diversity in public opinion (McCombs and Zhu 1995), and another on Web site browsing by and across domains (Li, Liechty et al. 2002). They offer somewhat appealing models for our question. Most belong to a class called "multiple count data" (Aitchinson and Ho 1989). In plain language: why did this site list four key issues under food security, but only one under health, etc., and how does the distribution over all sites relate to factors such as target group, etc.?

Practically, disasters have different effects for different communities, and when an effect exceeds a recognition threshold, it may become a "key issue" in contact with outsiders. Whether the key issue will actually be recorded as such, or not, depends on a number of factors that may have little to do with the disaster itself. The agenda capacity of the encounter, the coherence of community opinion, and the endeavor to secure attention for the most pressing problems will truncate the number of issues distinctly recorded.

Just as importantly, the assessment teams wield power as to what goes into the recording space. They exert it following their own cognitive schemes. Plausibly, low-intensity messages from the communities get recorded in one item, or not at all. As the intensity increases, more categories are considered and eventually used. At this level, they often are substituted for each other - e.g., of two candidates, "food availability" and "food accessibility", only one is used, but which it is can be rather accidental. At higher levels, more differentiated notes are created, resulting in more distinct issues.

This proposition can in part be tested. We expect that if a community ranks a sector as a priority, the number of key issues recorded under it will tend to be higher. This is indeed the case, although the difference is modest (see the table on the next page).

There is another problem with "key issues" at which we have already hinted. There are too many of them. At data entry, Yemen kept 59 distinct issues. The analysis needs to reduce this complexity. About this we will further speak in the recoding section.

The consequence is that issue rankings by frequency are not robust<sup>3</sup>, particularly if the medium-intensity issues are numerous. While that is a statistical problem, psychologically it shows up in the grade inflation that transpires from the severity ratings. To this we now turn.

---

<sup>3</sup> This is not a problem specific to needs assessments, and even less so to the Yemen. "Ranks are notoriously uncertain", remarks Tunaru (2002: 226), in a different context of multiple count data.

**Table 8: Key issues recorded, by sector and sector priority**

Sector	Sector is a priority for this community		
	No	Yes	Total
	Mean no. key issues recorded		
	Sites		
Education	3.43	4.67	3.60
	37	6	43
Food Security	2.56	2.29	2.35
	9	34	43
Health	2.95	3.35	3.16
	20	23	43
Livelihood	2.92	3.18	3.02
	26	17	43
Protection	3.00	N/A	3.00
	43	0	43
Return	1.56	N/A	1.56
	43	0	43
Shelter	3.14	3.13	3.14
	28	15	43
WASH	3.46	4.60	4.26
	13	30	43
<b>Total</b>	<b>2.80</b>	<b>3.38</b>	<b>3.01</b>
[Sites X sectors]	219	125	344

### Severity ratings

The severity ratings for key issues follow from the green to red color scheme. This tool seems to be easily understood and consistently filled in. In the Bolivia and Yemen templates, the scoring formula was built in, prompting "1" for green, up to "4" for red.

This is ordinal data. However, the severity scores represent ratings, not rankings. The scoring is independent among issues, at least formally. There is no first, second, etc. priority issue, as there is for priority sectors. At each level, arbitrarily many issues can be tied. The previous sidebar already talked about this. Here we address additional aspects.

The heat maps of the Yemen report are reportedly based on the mean, over the communities in point, of severity scores, given the issue. This is incorrect even if widely practiced and sanctified by some statisticians close to the participatory appraisal movement (Abeyasekera Undated). The scores do not have an interval scale basis.

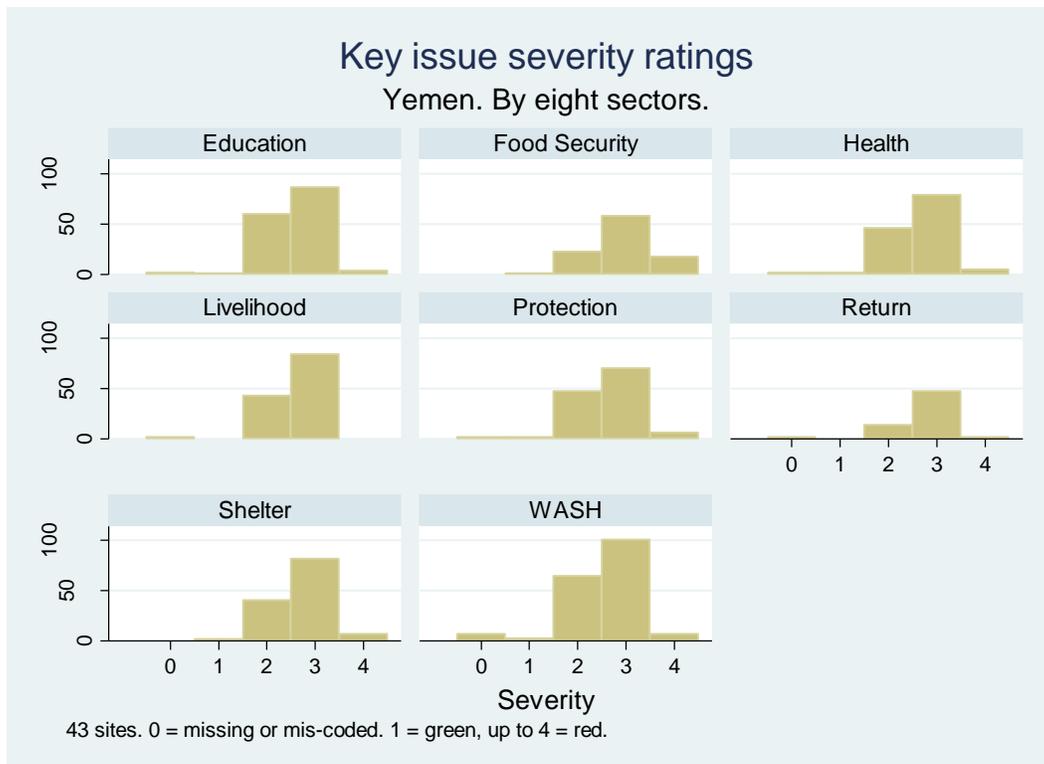
The error is probably due to an equally widespread misconception that we are dealing with Likert-type scales (Wikipedia 2011c). That is not the case. Likert scales are scores summed over items that arguably measure attitudes towards a common object.

By contrast, issue frequencies are over subjects (the sites). The Likert data format does not mean that a scale by this name is automatically formed. In theory, there could be other models producing valid interval-level scores, such as the polytomous Rasch model (Harwell and Gatti 2001). Not only are "key issues" not suitable for them (they are not normal test items), but such models are out of reach for most data managers.

What can be done in this situation?

In principle, both the maximum and the median are meaningful statistics of ordinal ratings. However, this is frustrated by: a. the tendency, the larger the summarized set, for the observed maximum to be the highest design value; in other words, almost everything in the heat map turns red. b. as for medians, Excel does not admit them in Pivot tables, and the de-facto grade inflation makes them meaningless; in other words, almost everything turns orange.

**Figure 2: Key issue severity ratings, Yemen, by sector**



Despite their limitation, there are productive uses for severity ratings:

1. The problem-level ratings can be used as an ordering scheme during visual inspection. We will demonstrate this below, with a Pivot table. Later, the key issues will be recoded, into smaller category sets that compress tables, graphs, and text in the report.
2. The sectors used in a given assessment - eight in Yemen - form a defined, closed set. Severity was rated also for each site-sector combination for which at least one issue was recorded. For those without noted issues, a normal situation regarding this sector and this community may be assumed. These ratings may be recoded to 1 (green). Then,
  - a. the distribution of scores, by sectors or on other dimensions, for synthesis-level records can be graphed out
  - b. Borda counts are still not feasible (the scores are not truly ranked). But it is legitimate to define a severity index as the number of sectors for which a site was rated severely affected (red).

Such an index, or its representation in a site X sector red/black heat map, would be more informative than the use of the non-category "cross-cutting".

## Visual inspection

As we continue the analysis work, visual inspection of tables remains the preferred method for detecting patterns, errors, and leads into new promising terrain. Tables force attention to detail. What exactly is the setup? Which category values were expected, but did not appear because there are no observations? Which disappear when we filter? Which become outliers when we filter because most of the others almost disappear? Etc.

Charts are particularly good at driving home the key message of a distribution or association, but they do not encourage the same intense attention to detail as tables do. Also, in Excel, Pivot tables are rapidly reconfigured; charts take longer.

This discipline demands economy and triage also in table production. By this, we mean both the number of tables to be retained for further work and ultimately for the report, and the complexity of every table. As regards key issues and severity ratings, the analyst may benefit from venturing from smaller sets to larger in several steps.

### Smallest circle

We inspect key issue data beginning with a small segment. Our hope is that we can rapidly form hypotheses of how urgent needs are distributed by

1. initially tabulating problems that were recorded in priority sectors and with the highest severity score (red = 4)
2. gradually expanding the list by including lower scores and/or non-priority sectors.

This pivot table filters to: 1. problem-level records, 2. in priority sectors, 3. of severity score 4. The row factor hierarchy places "Sector" above "Problem"; the column factors are: "IsPrioritySector" above "Severity". There is only one informative numeric column.

**Table 9: Pivot table of key issues, by sector, filtered to priority sectors and highest severity score**

Keep_record	Yes			
Site_level_record	No			
Synthesis_record	No			
<b>Count of RecordNo</b>		<b>IsPrioritySector</b>	<b>Severity</b>	
<b>Sector</b>	<b>Problem</b>	<b>4</b>	<b>1 1 Total</b>	<b>Grand Total</b>
<input type="checkbox"/> <b>Food Security</b>	Food accessibility	7	7	7
	Food availability	8	8	8
<b>Food Security Total</b>		<b>15</b>	<b>15</b>	<b>15</b>
<input type="checkbox"/> <b>Health</b>		0	2	2
	Health Status and Risks	4	4	4
	Malnutrition	1	1	1
<b>Health Total</b>		<b>7</b>	<b>7</b>	<b>7</b>
<input type="checkbox"/> <b>WASH</b>	Water supply/management	1	1	1
	Waterborn diseases	5	5	5
<b>WASH Total</b>		<b>6</b>	<b>6</b>	<b>6</b>
<b>Grand Total</b>		<b>28</b>	<b>28</b>	<b>28</b>

We find that only three sectors make it to this most exclusive list. Each shelters two key issues. Two health sector records must be defective, "0" appears where an issue category was expected. Two issues - "malnutrition" and "water supply/management" - meet the criteria - part of a priority sector, highest severity - in one community only.

Do these problems cluster in the same site, or for the same target groups? Since the number of records is relatively small (28), we may find it worthwhile inspecting the subset of the data table manually. More conveniently, we work with the Pivot table wizard, dragging those fields - one at a time - between "Sector" and "Problem".

When we do so, we find that (output omitted here) that there is little clustering of issues by site. Only in two sites were there both food availability and accessibility recorded. One site is on record for two health-related issues. In the WASH sector, every instance of an issue is from a different site. For target groups, the picture is slightly different; severe food accessibility issues plague IDP communities only.

### A wider set

We now widen the circle by also including problems scored 3 (orange) for key issues. We keep the priority sector restriction.

This makes the list considerably longer. It is less wieldy to quick inspection. Making two columns (for severity score 3 and 4) temporarily invisible facilitates inspection. We now have issues listed in five sectors. This screenshot displays the issues and their frequencies for two sectors.

**Table 10: Pivot table of key issues, enlarged to severity score 3 and 4 (segment)**

Keep_record	Yes	
Site_level_record	No	
Synthesis_record	No	
<b>Count of RecordNo</b>		
<b>Sector</b>	<b>Problem</b>	<b>1 Total</b>
<input type="checkbox"/> <b>Education</b>	Access and learning environment	3
	Child labour	1
	Education material/uniforms	6
	Enrollment of girls	3
	Teachers and other education personnel	1
<b>Education Total</b>		<b>14</b>
<input type="checkbox"/> <b>Food Security</b>		0
	Food accessibility	14
	Food availability	28
	Food diversity	12
	Food utilisation	3
<b>Food Security Total</b>		<b>58</b>

Frequencies of issues within given sectors differ more keenly. We begin to see, at least in part, how some of the rarer categories might be combined. Could the rarely used "food utilization" be folded into "food diversity", perhaps under a recoded term "Food, quality"?

We may also want to make a note of some issues that the teams recorded surprisingly rarely. There was only one community speaking of child labor at this intensity and with education as a priority. Is there more to it, but it is not considered serious? Will the issue return under different sectors? Those kinds of questions can be answered with greater confidence once when we inspect the unrestricted issue set. Right now, with these restrictions, we find it easier to come up with interesting questions in the first place.

The analyst, from this point onward, may also wish to make notes within a first draft outline of the report. If, as one would hope, the analysis is done by a team, some leading ideas, reminders, etc. may go up on sticky note boards or in shared digital documents, all subject to revision.

#### **Wider and narrower issue sets**

In similar manner, it is productive to inspect the key issues when we remove the restriction on priority sectors, but restrict issue ratings to "severe". In the Yemen table, this will bring up, among others, six different protection issues. Each was recorded at one site only. We may speculate that protection issues are highly variable and not enough focused for the communities to consider any one as particularly serious, or to consider the entire protection realm as a priority sector. The protection needs may be serious, but - in what the teams recorded! - they take a lesser seat behind other needs. Again, this is a

more or less arbitrary first reading of a data subset, meant to stimulate detection and discussion of patterns while visually inspecting a manageable subset of the data.

### A doubling at each step

The number of distinct key issue formulations (including defective ones like "0") as a function of filtering in the Pivot table is given in this table. Every time the bar is lowered, the number roughly doubles, from 8 to 17 to 33 to 58. We add information on sector and issues of lesser importance, and work out more complex interpretations.

**Table 11: Gradual expansion of key issue inspection**

Distinct issue categories		
Issue severity	Priority sectors	
	Sector is a priority for this community	All sectors
4	8	17
3 - 4	33	53
No restriction	35	58

The arrows suggest that - with data like Yemen's - it may be efficient to proceed with the visual inspection through four successively wider sets of issue records. The other two possible sets of restrictions do not promise analytic gain.

### Back from figures to text

Regardless of all the figurework, analysts who deal with multiple response data as variegated as these key issues will need to look at the text closely. What do these entries say? For example: How is "health resources and services availability" different from "health system performance"? Does anyone trust that the difference between those two was consistently handled in the coding of what the key informants had to say? In theory, the analyst can go back and consult the detailed notes that the teams made of their key informant interviews or community group discussions. Practically, time pressure and language barriers limit the effort to few and selective hardcopy inspections after the data entry has been completed.

Looking at numbers alone does not enlighten; the distinctions among issues matter first. The analyst needs to evaluate the meaning of text elements no fewer than three times:

1. Initially when the first categorization happens (what goes into the drop-down menus?).
2. Again during first analysis when issue frequencies and severity patterns show up in the Pivot tables.

The tables are fast to reconfigure, making visual inspection efficient. This must not tempt us to multiply output for the final report. Rather, we start from the tip of the iceberg, successively diving to the larger parts hidden under water, taking advantage of the

presence of field supervisors or translators who are still with us to help interpret candidate findings. These persons know the meanings of categories behind the numbers.

3. The third look is due when this discovery phase is over. There are still too many issue categories. They must be reduced through recoding.

The original categories will continue to play a role - some selectively in illustrations, most or all in more detailed tables in report appendices. Recoding presupposes, as in all qualitative research, that we 1. either understand what the original categories mean, or 2. if some meanings are vague, how we can absorb them in one or several broader categories that are still useful.

Before we move to recoding, we have one more use of severity ratings to discuss.

### Severity ratings in "synthesis" records

The Yemen assessment let key informants and community groups enumerate, and rate the severity of, problems by sector. When the conversation within a given sector was drawing to an end, the team would prompt the participants to offer a severity rating for the entire sector.

Such a rating, in terms of data relationships, is an attribute of the site-sector combination. Its numerical value is lodged in the "synthesis" record, in the same field that holds the ratings for the key issues. The value is established from the subjective judgments of the participants, not as a formula result. Issues of averaging key issue-level ratings therefore do not arise.

In the Yemen case, 5 percent of all recorded key issues were rated as severe. More than double that proportion, 11 percent, of the site-sector syntheses reached that grade. This table is the result of two small Pivot tables (and of the percentages calculated outside).

**Table 12: Severity ratings of key issues vs. those of synthesis records**

Severity	Key issues		Syntheses	
	Total	Percent	Total	Percent
1	11	1%	0	0%
2	340	32%	77	28%
3	646	61%	164	60%
4	57	5%	31	11%
<b>Total</b>	<b>1,054</b>	<b>100%</b>	<b>272</b>	<b>100%</b>

We compare this with the priorities that the same communities identified during the assessments when asked about their top-three priority sectors.

**Table 13: Agreement between severity ratings and priority judgments regarding sectors**

Severity ratings for the entire sector	Is this sector a priority for the community?		Total
	No	Yes	
2	58	19	77
3	105	59	164
4	12	19	31
<b>Total</b>	<b>175</b>	<b>97</b>	<b>272</b>

Note: In different phases of the interviews / discussions, the same persons, regarding the same sector, ranked it as a priority sector (1st, 2nd, 3rd or none) and rated the totality of key issues they had enumerated for it ("Synthesis"). The table shows the coherence of 272 site-sector combinations for which both propositions were recorded.

The agreement between priority sectors and synthesis judgments is minor. Only in a fifth of all situations where the local participants ranked a sector as one of three priorities did they also grant a "severe" grade (19 out of 97). And, only for two thirds of the "severe issues" (19 out of 31) did they indicate that these sectors were among their priorities.

Whether one should consider the low agreement disappointing is a philosophical question. Interviews and group discussions involve dynamic learning processes on both sides. The participants are certainly entitled to second-guess their sector priorities when asked to evaluate them in a subsequent, different format (the "synthesis") as long as both propositions are faithfully reported and available for comparison.

Practically, key findings are faster to explore by starting with the 19 cases in agreement, and next with the 31 synthesis records rated "severe". Inspecting and describing the 97 priority sector instances will plausibly take more effort.

## Recoding

### Verbose entries vs. drop-down lists

Survey response can be recorded in free-text or in standardized (categorized, pre-coded) version<sup>4</sup>. Depending on this form data tables differ in important respects.

In Bolivia, problem formulations were not standardized at entry. In Yemen, they were, in lists that controlled the drop-down menus. Our demo workbook follows the Yemen case. Since verbose data handling as in Bolivia may occur in future assessments, we need to briefly discuss the use of recoding in this scenario.

---

<sup>4</sup> This concerns recording, not the form of the interview question (open vs. closed). Response to open questions can be recorded in both ways.

A reminder: In the Excel workbook template, the category sets are kept as list in separate sheets, i.e. in the sector-wise lists. It is in these lists that we build recoding schemes. They are the lookup tables for the formulas that do the actual recoding in the data table.

### **Recoding verbose entries**

In a slow-moving survey, verbose entries would be preferable because they preserve more of what the key informants actually said. They would enable a slow, iterative process of interpretation. Rapid needs assessments do not have this luxury. In all but very small datasets, verbose data threatens the analysis process with delays or eventually disregard for portions of the data. The threat has to be disarmed with the help of recoding.

Given the verbose problem and recommendation formulations, the data entry proceeds without drop-down menus for these variables. This was clearly the case in Bolivia. Therefore, even assuming that our data table template is being used, the sector-wise lists have not yet been populated by the time when analysis begins with the finished data table.

How should the recoding be done in this situation?

Since most of the verbose entries are unique, we may be tempted to recode manually in the data table itself. This would be short-sighted.

Rather, because a first coding was not done at data entry, there may be two stages required to come up with small enough category sets. The analyst must be able to visually focus on entries in one sector. She needs to be able to sort them by her tentative recodes, unencumbered by other sector categories. This is possible by

- building recoding schemes in tables separated by sector, then
- letting the built-in lookup mechanism do the actual recoding in the data table.

This takes several steps. For key issues, these are:

1. At first, we build, in a new sheet, a Pivot table yielding simple frequency table, filtered to one sector at a time, of the verbose key issue entries.
2. We copy - paste (values only) the tables into column 1 and 2 of the concerned sector sheets. The frequencies will be trivial (1 in most cases).
3. We then "recode a first time"(more correctly: "code for the first time"), by trying out problem categories that each cover a number of the verbose entries. The recodes are in column 3. After the first attempt, we sort the table on this column. We do this in all sector sheets.
4. [Optional:] If the table in a sector sheet is long - i.e., there are numerous distinct verbose entries - and we still have many distinct recoded categories, a small frequency table for these, further down in the same sheet - may help us.
5. We keep modifying the recodes and sort and inspect repeatedly, until we have a satisfactory set of categories for grouping the key issues inside the sector.

The same procedure will be suitable for recoding the recommended interventions, in column 6 - 8 of the sector sheets.

### Recoding drop-down menu options

In the Yemen-type situation, the drop-down menu options are recoded. The challenges are different.

Some of the work is easier: The frequencies are more informative. Rarely used options are easily spotted and can be collapsed into other categories if this makes sense. The absorbing category and the absorbed one(s) may be given a new name in the recoding column.

Difficulties, however, may arise when the drop-down menu options carry little information. The re-coding may then do violence to the initial intent that prompted the drop-down options used in the data entry. However, not to recode is impractical. Moreover, outsiders faced with tables that use the original options would not always understand what they mean.

We illustrate this with the recodes that we used for the shelter sector problems in Yemen.

**Table 14: Example of a recoding scheme - key issues in the shelter sector**

Problems	Frequency	Problems_Recoded
Shelter Security/Condition	60	Safety and security
NFI status	40	Financial
Heating/Cooking	17	Financial
Disaster risk	5	Safety and security
Expensive Rent	13	Financial

Except for the last option ("expensive rent"), none of the initial formulations lets us clearly understand what the problems were. For example, what kinds of disasters were looming? Fires in crowded camps? Epidemics? In such a situation, it will be daring, but ultimately more productive, to recode into a small number of broad categories with which the initial options share a plausible overlapping rapport. Here we assume that in three of the five problem types recorded, poverty is a common cause. In the two other problems, we assume strong safety and security connotations.

The information loss in this recoding is obvious. For example, shelters may be in poor condition for reasons that have nothing to do with safety and security.

Any recoding that reduces the number of categories entails some information loss. The sidebar further below discusses the extent of such losses. Here we will first see how tables using the recoded problems compare to tables discussed earlier.

### Counting recoded problems and their relative severity

A potentially (sic!) productive Pivot table to start with may be configured like this:

**Table 15: Key issues, recoded, by sector and severity**

Keep_record	Yes						
Site_level_record	No						
IsPrioritySector	(All)						
Synthesis_record	No						
<b>Count of RecordNo</b>		<b>Severity</b>					
<b>Sector</b>	<b>Problem_recoded</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>Grand Total</b>	
Education	Education, access	1	41	69	4	115	
	Education, delivery		18	15		33	
Food Security	Food, effective access		11	37	18	66	
	Food, quality	1	12	21		34	
Health	Health care capacity	1	16	36		53	
	Health condition	1	21	40	5	67	
	Human resources		9	3		12	
Livelihood	Financial		27	79		106	
	Human resources		15	4		19	
Protection	Conflict, crime and violence	1	35	59	5	100	
	Safety and security		4	8	1	13	
Return	Financial		5	22		27	
	Safety and security		9	24	2	35	
Shelter	Financial	1	26	40	1	68	
	Safety and security	1	15	42	6	64	
WASH	Hygiene		5	3		8	
	Sanitation	1	16	14		31	
	Water	2	42	82	7	133	
<b>Grand Total</b>		<b>10</b>	<b>327</b>	<b>598</b>	<b>49</b>	<b>984</b>	

The cell values express recoded problem instances over all sites. As such, they are not to be taken for the number of sites that expressed some key issue at those levels of severity. Looking at the figures, we may think that "effective access to food" represents one of the most urgent, if not the most urgent, needs in the entire affected population. Yet, as it is, this table is not compelling. Key informants at different sites, interviewed by different teams with different inclination to record problems, may have inflated or deflated the diversity and severity of problems.

A filter is needed. We require that the recoded problem instance be in a sector that the concerned community listed as one of three priority sectors.

This shrinks the list to fewer problems.

**Table 16: Same table, with problems restricted to priority sectors**

Keep_record	Yes	<input type="checkbox"/>					
Site_level_record	No	<input type="checkbox"/>					
IsPrioritySector	1	<input type="checkbox"/>					
Synthesis_record	No	<input type="checkbox"/>					
<b>Count of RecordNo</b>			<b>Severity</b> <input type="checkbox"/>				
<b>Sector</b>	<b>Problem_recoded</b>	<input type="checkbox"/>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>Grand Total</b>
Education	Education, access			7	13		20
	Education, delivery			4	1		5
Food Security	Food, effective access			9	27	15	51
	Food, quality		1	10	15		26
Health	Health care capacity			5	25		30
	Health condition			10	23	5	38
	Human resources			4	1		5
Livelihood	Financial			7	34		41
	Human resources			9	2		11
WASH	Hygiene			4	3		7
	Sanitation		1	11	10		22
	Water		1	29	62	6	98
<b>Grand Total</b>			<b>3</b>	<b>109</b>	<b>216</b>	<b>26</b>	<b>354</b>

We now need to set rules as to how to signal priority problems. These rules can vary from assessment to assessment, and from analyst to analyst, as long as they are explicit and offer some rationale.

### Prudence and fairness

With this Yemen data, we might want to

1. first adopt a rule of **prudence**, to say that any problem recorded as "severe" more than once *in this filtered table* should be highlighted red for attention.
2. next adopt a rule of **fairness** for the distinction between orange (serious risk) and yellow (more data needed). For example, *in the unfiltered table*, any problem would be marked orange if it was not yet marked red and either had at least two severe instances or more than 10 instances of serious risk (10 being a quarter of the number of sites [43]).
3. The remainder of problems would be highlighted yellow.

What is the rationale for these rules?

The first rule limits consideration to problems that arose in sectors of priority. Within these, if more than one instance was recorded, we err on the side of prudence and raise attention to a level as though the key issue were severe everywhere.

The second rule takes account of the fact that the absolute number of problems rated as "of concern" does not correspond to the number of communities so affected. Yet, above a certain threshold, any problem should be considered as being of concern (orange). Also, if in non-priority sectors more than one outlier of "severe" appears, similar concern should be indicated.

The third rule simply says that no recoded key issue be left in the green category. This rule would be broken in the (unlikely) event that some recoded issue absorbed only original issues all of which had been rated as "no major problem" (i.e., originally all of them were green).

### Greatly reduced complexity

The resulting table, in slightly edited form, would then look like:

**Table 17: Key issues, recoded and color-marked for attention**

Key issues by sector						
Sector	Issue	Severity				All problems
		No major problem	Lack of data	Of concern	Severe	
Education	Education, access	1	41	69	4	115
	Education, delivery		18	15		33
Food Security	Food, effective access		11	37	18	66
	Food, quality	1	12	21		34
Health	Health care capacity	1	16	36		53
	Health condition	1	21	40	5	67
	Human resources		9	3		12
Livelihood	Financial		27	79		106
	Human resources		15	4		19
Protection	Conflict, crime and violence	1	35	59	5	100
	Safety and security		4	8	1	13
Return	Financial		5	22		27
	Safety and security		9	24	2	35
Shelter	Financial	1	26	40	1	68
	Safety and security	1	15	42	6	64
WASH	Hygiene		5	3		8
	Sanitation	1	16	14		31
	Water	2	42	82	7	133
<b>All problems</b>		<b>10</b>	<b>327</b>	<b>598</b>	<b>49</b>	<b>984</b>

Instances of key issues mentioned by communities, by degree of severity.  
Severity rating of issue determined also by the the sectors that had priority for given communities.

The result is a table of eight sectors and 18 key issues, marked in three colors. The figures to the right are understood with the help of the two footnotes. Substantively, this table is the equivalent of the table in the summary of this note, except for the figures.

The complexity of this output is considerably less than that of any table ordering the 59 initial problem types. It is the result of recoding and of the rule-based interpretation process.

### Where does the interpretation happen?

In recapitulating what we just did in building the last table, it is important to recall that this interpretation takes place *during* table production. We order, group and interpret key issues together with their severity as rated by the communities. What the analyst adds is

the rules of interpretation. In the Yemen case, with its sector priorities as well as issue severity ratings, we *invented rules* of prudence and fairness. In other assessment configuration, different rules might be appropriate.

Conversely, in tables summarizing synthesis records, severity has been rated already *before* the data entry. This is an interpretation process that the teams prompt the key informants and community groups to do in consideration of all the problems they raised minutes earlier.

Thus, what we are dealing with here is a hierarchy of interpretation processes:

- Local participants interpret questions posed by team members, review their own understanding of the disaster consequences, and respond with enumerations of key issues as well as with a severity rating for each issue.
- The participants are then helped to pass a "synthesis" rating on all problems enumerated under that sector.
- The assessment team's interpretation effort in turn is focused on determining whether the elicited problems fall under the concerned sector and the severity ratings for them and for the synthesis reasonably agree.
- Finally, the analyst adds his own layer of interpretation, as described above.

While the fact that the data resulted from multiple interpretations is irrelevant for the mechanics of the Pivot tables, we ought to always remain aware of the extreme artificiality of the processes from site visit to final table. This is evident when we consider alternatives. For example, it would be just as plausible to have the assessment teams give the "synthesis" ratings after the site visit, instead of having the local participants review the ratings of all the concerned key issues and, by some undocumented aggregation logic, compact them into one figure.

### **Transparency in recoding**

Most consumers of the assessments will not have access to data tables in spreadsheets. Those few who bother to open an Excel file will not want to work through arcane recoding schemes. Regardless of the mechanics, the readers must be able to see how initial problem types were renamed and grouped. The recoding should thus be documented, probably in an appendix to the assessment report.

This too can be done, and easily so, in a Pivot table. This screenshot shows the set-up. As with all such tables, it needs editing. While we banish them to the appendix, particular segments may be used in the main body if detail is required to make a particular point.

**Table 18: Documenting how the recoding was done (segment)**

Keep_record	Yes		
Site_level_record	No		
IsPrioritySector	(All)		
Synthesis_record	No		
<b>Count of RecordNo</b>			
<b>Sector</b>	<b>Problem_recoded</b>	<b>Problem</b>	<b>Total</b>
Education	Education, access	Access and learning environment	47
		Child labour	4
		Community participation	1
		Education material/uniforms	40
		Enrollment of girls	24
	Education, delivery	Education policy/Coordination	2
		Teachers and other education personnel	27
		Teaching and learning	5
Food Security	Food, effective access	Food accessibility	27
		Food availability	40
	Food, quality	Food diversity	30
		Food utilisation	4

In the edited version, recoded problems could be called "Key issues" and problems "Specific issues". It is important to let the reader understand that the figures under "Total" mean instances of specific issues recorded in all assessed communities. They are not the numbers of communities that raised the issues<sup>5</sup>.

### Other tables, and tabulating recommendations

Other tables, such as those obtained by adding target groups as a column variable, follow the same spirit of rule-based interpretation. The thresholds may have to be adapted to the smaller numbers that results when the sample is broken down on more dimensions.

The recoding of recommendations pursues the same objective - the reduction in the number of categories and their grouping in meaningful new categories. Recommendations are not severity-rated.

A complication arises in relating recommendations to key issues once both have been recoded. The sector-problem-recommendation combinations actually realized will be far fewer than before recoding. In the Yemen data, they drop from 146 to 42. This achieves the desired reduction of complexity.

However, this comes at the expense of transparency. In the workbook, it may be easy to create a hierarchical Pivot table listing, for each recoded combination, the original

<sup>5</sup> The calculation of this statistic is more involved. In Excel, a community identifier would be used as a column label in this Pivot table. To the right of this expanded table, for the same number of communities dummy variables would be created, with an IF function returning 1 if the community had a positive number of instances of the given issue, and else 0. The sum of these dummies is the number of communities having raised the issue. For an editable version, the result would have to be copied and pasted (values) to the left.

recommendations that it embraces. Documenting this in an assessment report would be tedious, possibly requiring a six-column table in landscape format.

What is practical? Probably only the cross-tabulation, one sector at a time, of recoded key issues with recoded recommendations. This table demonstrates the set-up for one sector - food security.

**Table 19: Cross-tabulating recoded key issues and recommendation - food security**

Keep_record	Yes				
Site_level_record	No				
Synthesis_record	No				
Sector	Food Security				
<b>Count of RecordNo</b>	<b>Recommend_</b>				
<b>Problem_recoded</b>	<b>Advocacy and awareness</b>	<b>Economic support</b>	<b>Training</b>	<b>Grand Total</b>	
Food, effective access	10	57		67	
Food, quality	1	32	1	34	
<b>Grand Total</b>	<b>11</b>	<b>89</b>	<b>1</b>	<b>101</b>	

Note: Read "Recommend\_Recoded" for the column variable.

What if more detailed investigation is desired? Again, the Pivot table lets us do this rapidly and conveniently, either by

- double-clicking the cell and inspecting the table created, or by
- dragging the original key issue and/or recommendation variables into the Pivot wizard.

The first method lets us see the subset in the context of any other variables of concern at this point. The second is efficient for comparison among sub-groups, but displays only issues and recommendations.

### **[Sidebar:] How much information is lost in recoding?**

To answer this question, we need a measure expressing information diversity. In studies of public opinion, one often used for thematic diversity (McCombs and Zhu 1995) is Shannon's entropy measure  $H = - \sum_i p(x_i) \log_b p(x_i)$ , where  $p(x_i)$  is the probability (or observed relative frequency) of the  $i$ -th item in a set of  $n$  items, whose probabilities or relative frequencies sum to 1 (Shannon 1948; Wikipedia 2011b). This example calculates the entropies of the initial WASH sector key issues as well as of the recoded ones, using natural logarithms. The calculation can be found also in the WASH list of the demo workbook.

**Table 20: Calculating the entropy of a key issue distribution**

Problems	Frequency	Problems_Recoded
Water supply/management	40	Water
Water sources	28	Water
WASH NFIs status	31	Water
Sanitation/Excreta disposal	32	Sanitation
Hygiene Practice	9	Hygiene
Water-borne diseases	11	Water
Water quality	28	Water

Problems	Proportion (p)	-p * ln(p)
Water supply/management	22%	0.335
Water sources	16%	0.290
WASH NFIs status	17%	0.304
Sanitation/Excreta disposal	18%	0.308
Hygiene Practice	5%	0.150
Water-borne diseases	6%	0.171
Water quality	16%	0.290
Sum	100%	1.848

Problems, recoded	Proportion (p)	-p * ln(p)
Hygiene	5%	0.150
Sanitation	18%	0.308
Water	77%	0.201
Sum	100%	0.659

<b>Info diversity loss (1 - 0.66 /1.85)</b>	<b>64%</b>
---	------------

The information diversity loss incurred by recoding, within this sector, is about 64%. This is larger than the relative reduction in the number of categories, which is  $(7 - 3) / 7 = 57\%$ . The merging of water-related categories created one highly used water category (77% of all key issue instances in WASH). It contributes little to the entropy of the distribution of recoded key issues. This suggests that the recoding was too radical.

A similar loss of information diversity results in the recoding of issues over all the eight sectors<sup>6</sup>. In strict information-theoretic terms, the recoded-issue distribution, in our case here, is less efficient. However in terms of *cognitive* efficiency - the effort required on the part of the assessment consumer to assimilate the information - there is a gain of about 20 percent. This

<sup>6</sup> Approx. 63%. The calculation is more complex; it requires the conditional entropy  $H(\text{key issues} / \text{sectors}) = H(\text{key issues}, \text{sectors}) - H(\text{sectors})$  for original as well as recoded issues. The efficiency of the recoded-issue distribution is lower because the change in denominators  $[\ln(\text{number of categories})]$  is not enough to offset the change in entropy. The cognitive efficiency, by contrast, divides the entropy by the absolute number of categories.

seems tiny, but keep in mind that our recoding is reckless for didactic reasons, slashing categories by a factor of almost 3.5.

That is about efficiency, not cognitive *effectiveness*. If assessment consumers are to take home strong messages, tables must be small and highlight drastic differences.

Finally, any philosophizing about recoding must keep its sights on a humbling basic fact. We suffer the greatest information loss long before recoding - during interviews, questionnaire notes, translation and categorizing at data entry. The community may have said: "See, over there, those puddles of stagnant water? Mosquitoes are plenty hereabouts. Several of us have come down with malaria." This may be noted down as "Stagnant water - malaria." It enters the data table as a drop-down menu option named "Water-borne disease", and finally gets recoded, alongside other issues, as "Water". Recoding can, of course, be smarter than in this example, but loss of information already beforehand is unavoidable. It is inherent in all categorizing:

*"To categorize is to render discriminably different things equivalent, to group the objects and events and people around us into classes, and to respond to them in terms of their class membership rather than their uniqueness"* (Bruner, Goodnow et al. 1972).

---

## **When does analysis end?**

The substantive analyses conclude when

- the analyst feels that the further production of tables detailing combinations of two or more among these: *region, target group, sector, recoded problems, recommended interventions* no longer yields significant insights
- the analyst has written down the outlines of the necessary interpretations (i.e. is ready to mete out death by PowerPoint)
- key findings and interpretations have been deliberated, in the assessment team and, as far as practical, with the stakeholders.

"The analyst", of course, may be an individual, a group of persons, or a group reinforced remotely. A well-known risk is the loss of contact with team members knowledgeable of the local context ("what the data mean") who move on before the analyst has had a chance to cross-check unusual patterns or anomalies with them.

### **Produce a lot of tables, discard most**

The end of analysis is not the end of assessment work. But there is a strong need to set a cut-off point, in order to leave time for editing and documenting. During the analysis process, tables will be produced and

- looked at once and then discarded (probably most)
- saved as an information copy in a research journal, with brief notes (some)
- saved in workbooks as key finding supports (10 - 20), or
- saved in workbooks for report appendices or other documentation (numbers will depend on regions, target groups, sectors, etc.).

Essential findings can usually be discerned in tables. Analytic statistics are generally not required, and are often not feasible given the nature of the sample. The production of charts can wait until the team has determined what the essentials are.

### **Accelerate maps, delay charts**

The case of maps is different: tables reveal spatial clustering only partially even when an administrative hierarchy is used. Particularly when the team member producing the tables is not also the GIS expert, requests for maps may have to be made at several points in the analysis process. This calls for close coordination, both teamwork-wise and for agreement between key messages from maps and those from the statistics. Having maps early in the analysis process stimulates hypotheses and aids interpretation.

### **Set a cut-off point for analysis**

Because contexts differ from assessment to assessment, a standard list of tables, graphs and maps to be produced is not practical. Depending on time pressure and deadlines, the cut-off point for analysis work comes sooner or later. There may be a case for drastic triage, as implied in the table on page 28. In the extreme, a timely, 3-page note limited to issues that the communities rated "severe" may be worth more than a 30-page report covering the whole gamut, but delayed and reduced to historic interest.

## **Outlook**

This note has confronted two questions:

- *What can be done validly with key issue-centered assessment data?*
- *What are the tools needed if assessment teams are to have a fighting chance under the tough Phase-2 deadlines?*

We limit the tool recommendations for the analysis phase to just two: Pivot tables, and recoding. Regarding validity, we show limits and strengths of various measures. We emphasize again the importance of eliciting priority sectors, which gives us a truly ranked measure as opposed to simple ratings.

### **Consider other approaches**

So far, so good. But at this point, we want to step back from the particular ways in which this note suggests to organize and analyze assessment data. We have noted the weak agreement that the current format, as used in Yemen, established between two measures - sector priorities and sector "synthesis" judgments. This lack of agreement suggests that one or the other or indeed both measures were not valid<sup>7</sup>. Or, if they were valid, then the learning processes during the meetings were so rapid and radical that the participants changed their evaluations between the first and the second measurements - in other words, they were not reliable.

---

<sup>7</sup> Technically, instead of "lack of agreement" between two measures, it would be more correct to say that the data did not support the theoretical implication that the preference order ("priority sectors") had for the distribution of severity judgments ("synthesis" ratings for the same sites and sectors).

Both possibilities motivate the search for different data collection approaches. These, in turn, will call for different data management and analysis templates. New approaches and formats can be considered that differ in increasing degrees from what was practiced in Yemen:

1. In a relatively minor change, the synthesis ratings would be given by the assessment teams, not the speakers for the assessed communities. This would relieve the meetings from having to synthesize a site-sector severity rating out of several key issue ratings.
2. More distinctly from the Yemen approach, while the local participants would still define priority sectors and would continue to express problems against all the sectors listed in the questionnaire, they would no longer be asked to produce severity ratings. Severity ratings would come only from the assessment teams and only in the sector syntheses.
3. In a radical departure from the setup followed so far, the topics and their evaluations by the community would be separated from those produced by the assessment teams.

How might this last option work? Key informants or small groups of community people would be stimulated to do a first screening of problems. To do so, they might respond to an abbreviated Hesper Scale-like battery of questions of the "Does this community have a problem with XYZ?"-type. An additional "Are there any other important problems?"-stimulus would secure attention for other potentially important topics. A minimum of three priorities would be established. Within these, specific problems and recommendations would be elaborated and taken note of in free-text fields. The questionnaire would be so designed that supplementary ("scoping") questions kicked in if certain particular problems had come up.

This procedure would not use a pre-defined list of sectors in talking to community people. For the assessment team, however, an observation schedule with a sector grid would organize their own chief areas of attention, problems definitions and severity ratings. The "areas of attention" would in part be derived from insights that the Phase-1 assessment offered. They could be written as a sector-subsector list. The subsector list would be purely ad-hoc for the particular disaster since there is no universally accepted one.

The separation between community-based evaluations (essentially the sector priorities) and the team's severity ratings (of issues organized by subsectors) would, of course, not guarantee strong agreement. But, at least, any disagreement could now be considered an expression of genuinely different perspectives, rather than of measurement error.

However, the weak-agreement problem would in part be shifted rather than resolved. The team evaluations would probably lead the data management, on the strength of its clear-cut sector-subsector structure (as opposed to the more open key issue elaborations by the communities). The inclusion of the community response, particularly the linkage between the community's priorities and the team's sectors/subsectors, would be challenging. It would call for recoding decisions, some of which might be highly arbitrary. And the time

that the teams would save while talking to communities (as compared to the current approach with severity ratings done by them) might be lost again in internal discussions.

### **Whatever the formats, the pressures remain**

There is no comfort zone in Phase 2; we deliberately use the word "triage" to encourage painful analytic choice. The fewer tables are retained, the more time the team (and stakeholders) will have to sort out what the apparent patterns in the data mean. Severity ratings are prone to abuse, to manipulations that are statistically illegal and to proliferating visuals that are suggestive, but conceal from the assessment consumer that they are anything but robust. Key issues may facilitate linkage with subsequent response planning, but the balance between preserving their diversity and reducing complexity is tricky. Together, these challenges call for analysts who impose soundness and restraint.

The alternatives that we sketched above have not yet been thought through for their consequences in data management and analysis. Even less have any detailed templates been developed for them. Neither the format used in Yemen nor any alternatives have been tested in the fire of a Phase-2 effort to deliver within two weeks.

It may well turn out that they cannot stand up to the pressure. Or that the stakeholders are in fact more liberal with the timelines and coordination needs that the assessment scripts stipulate. This would mitigate some of the trade-offs. One can imagine a core product to be made for on-time delivery, and a slower back office effort following up with more detailed, better edited documentation. The debate about suitable formats should go on for some time, and be enriched with the experience of some more assessments, before the concerned community of practitioners decides on its one preferred option.

## **References**

- Abeyasekera, S. (Undated). QUANTITATIVE ANALYSIS APPROACHES TO QUALITATIVE DATA: WHY, WHEN AND HOW Reading University of Reading.
- ACAPS (2011). Joint Rapid Assessment of the Northern Governorates of Yemen [9 October 2011]. Sana'a, Assessment Capacities Project (ACAPS), in collaboration with ADRA Yemen, CARE International, Save the Children, OXFAM, and Islamic Relief. Prepared for CARE International in Yemen.
- Aitchinson, J. and C. H. Ho (1989). "The multivariate Poisson-log normal distribution." *Biometrika* **76**(4): 643-653.
- Benini, A. (2011a). Data Management and Data Analysis - Key Issues in MIRA Phase 1 and 2. Washington DC and Geneva, ACAPS.
- Benini, A. (2011b). A template for managing data in needs assessments, centered on sites, sectors, problems, and severity of needs. Geneva, ACAPS.
- Borda, J. C. (1781). "Mémoire sur les élections au scrutin." *Histoire de l'Académie royale des sciences* **2**: 85.
- Bruner, J., J. Goodnow, et al. (1972). Categories and cognition. *Culture and cognition*. J. P. Spradley. New York, Chandler. 168-190.

- Garfield, R., C. Blake, et al. (2011). Common Needs Assessments and humanitarian action [HPN Network Paper No. 69, January 2011]. London, Overseas Development Institute (ODI).
- Global Education Cluster (2010). The Joint Education Needs Assessment Toolkit. Geneva, Education Cluster Unit, Save the Children.
- Harwell, M. R. and G. G. Gatti (2001). "Rescaling Ordinal Data to Interval Data in Educational Research." *Review of Educational Research* **71**(1): 105-131.
- IASC (2009a). INITIAL RAPID ASSESSMENT (IRA): FIELD ASSESSMENT FORM. Geneva, United Nations Inter-Agency Standing Committee.
- IASC (2009b). INITIAL RAPID ASSESSMENT (IRA): GUIDANCE NOTES. Geneva, United Nations Inter-Agency Standing Committee.
- Li, S., J. C. Liechty, et al. (2002) "Modeling category viewership of Web users with multivariate count models [Working paper no. 334]." Retrieved 28 October 2011, from <http://repository.cmu.edu/tepper/334>.
- McCombs, M. and J. H. Zhu (1995). "Capacity, diversity, and volatility of the public agenda." *Public Opinion Quarterly* **59**(4): 495-525.
- NATF (2011a). The Multi-Cluster/Sector Initial Rapid Assessment (MIRA) Approach. Process, Methodologies and Tools. Provisional Version as of 1 December 2011. Geneva, IASC Needs Assessment Task Force.
- NATF (2011b). Operational Guidance for Coordinated Assessments in Humanitarian Crises [Provisional version, 1st February 2011]. Geneva, UNOCHA and IASC Needs Assessment Task Force.
- Shannon, C. E. (1948). "A Mathematical Theory of Communication." *Bell System Technical Journal* **27**(July, October): 379–423, 623-656.
- Tunaru, R. (2002). "Hierarchical Bayesian models for multiple count data." *Austrian Journal of Statistics* **31**(3): 221-229.
- WHO and King's College London (2011). The Humanitarian Emergency Settings Perceived Needs Scale (HESPER): Manual with Scale. Geneva, World Health Organization.
- Wikipedia. (2011a). "Borda count." Retrieved 28 March 2011, from [http://en.wikipedia.org/wiki/Borda\\_count](http://en.wikipedia.org/wiki/Borda_count).
- Wikipedia. (2011b). "Entropy (information theory)." Retrieved 1st November 2011, from [http://en.wikipedia.org/wiki/Shannon\\_entropy](http://en.wikipedia.org/wiki/Shannon_entropy).
- Wikipedia. (2011c). "Likert scale." Retrieved 28 October 2011, from [http://en.wikipedia.org/wiki/Likert\\_scale](http://en.wikipedia.org/wiki/Likert_scale).

AB / 2 November 2011/ Revised 6 and 24 January 2012

The Excel demonstration workbooks are currently called:

For data management:

120105\_Assessment\_DataManagement\_Template.xlsx

Filled with the Yemen data, for analysis:

120124\_YemenData\_Analyses\_Demo.xlsx