LIS Operational Protocol P09 v 3 – Data Analysis

1.0 Purpose

1.1 The purpose of this protocol is to set guidelines for preparing and analysing the data from community surveys for the reporting phase of the LIS.

2.0 Overview

2.1 The data analysis portion of the landmine impact survey is essential for giving explanatory power to the outputs of the community survey process. For data analysis to function well, three related processes, outlined in this protocol, must occur. Firstly, HQ staff must modify and test query trains that convert data from the IMSMA into master tables in spreadsheets. Secondly, data must be analysed at the HQ staff level using basic descriptive statistics according to the needs of the final report. Thirdly, SAC staff may conduct more advanced statistical analysis, again according to the needs of the survey.

3.0 Responsibilities

2.1 HQ staff is responsible for ensuring that country-specific changes to the IMSMA field module data pool are compatible with Master Table query trains, and for carrying out basic descriptive statistical analysis for the LIS final report.

2.2 SAC is responsible for training HQ staff to recognise and correct problems with data conversion from IMSMA to Master Tables, and for conducting more advanced statistical analysis for the LIS final report.

4.0 Definitions

4.1 IMSMA: Information Management System for Mine Action; international standard database system for the LIS, technical and completion surveys and for other aspects of mine action.

4.2 Master table: In the context of LIS, a major grouping of data according to modules in the questionnaire. There are three Master Tables: for whole communities, for recent victims, and for danger areas.

4.3 Query: In the context of a database application, a set of instructions that allow for the export of data from the database into a spreadsheet table.

4.4 Query train: A set of queries that fit together to form a data series on spreadsheets

5.0 Instructions

5.1 Database conversion tasks

5.1.1 During training, SAC will familiarize HQ staff with the need for conversion of data from IMSMA to Master Tables and how this is done in database and spreadsheet applications. At least one person at survey headquarters must understand the nature of this process and the tasks involved. **[cf. 6.1]**

5.1.2 SAC, along with GICHD, will construct a series of relational queries **[cf. 7.1]** that find, convert and export data from the IMSMA database, where impact data is entered and stored, into Master Tables for victims, for SHAs, and for communities in spreadsheets.

5.1.3 HQ staff should, during the pilot tests, make sure that the query train template **[cf. 7.1]** developed by SAC is able to fully convert IMSMA-entered data into spreadsheets. If the survey makes use of IMSMA's user-defined variables, HQ staff should modify the template to include these variables in the query train output. **[cf. 6.2]**

5.1.4 Once the above procedure has been tested, and modified if necessary, and data from the field portion of the survey has been collected, HQ staff will export data from IMSMA into Master Tables in a spreadsheet format. HQ staff will document these tables, format them as needed, and add some calculated fields if necessary for the statistical analysis. SAC will modify the Master Tables as needed for export into a statistical application. All actors will document export of data from one application to another. **[cf. 6.3]**

5.2 Statistical Analysis at the Reporting Phase

5.2.1 HQ staff will conduct the basic descriptive statistical tests on the Master Tables that are required for the LIS final report, using statistical functions hardwired to spreadsheet applications. HQ staff may hire a national social scientist with data analysis skills for these tasks, who can also train field editors and conduct data monitoring in the field. HQ staff should set up a short training period for all staff involved in data management. **[cf. 6.4, 6.5]**

5.2.2 SAC will convert and export Master Tables on spreadsheets to statistical software, and produce more advanced statistical analysis, if no qualified statisticians are found in the survey country. SAC's statistical consultant will estimate the total affected population and the total number of affected communities, or verify a national consultant's estimation.

5.3 Case Studies

5.3.1 HQ staff may decide to include specific statistics of case studies in the report, to highlight some of the specific features of mine-affected communities. HQ staff should pick communities for statistical profiling that highlight key contrasts between variables. **[cf. 6.6]**

6.0 Rationale / Background

6.1 Master Tables form the basis of the descriptive analysis at the heart of a landmine impact report, so HQ staff should be comfortable with creating and evaluating them. **[cf. 5.1.2]**

6.2 This procedure should be executed by the GIS/database officer, with SAC assisting if necessary. Particular attention should be paid when setting up the values of country specific explosive devices. **[cf. 5.1.3]**

6.3 Some of the results of statistical tests may become variables that are entered back into the spreadsheets, to graph or use in pivot tables. GIS applications may also create data mapping exercises from the spreadsheets. **[cf. 5.1.4]**

6.4 Microsoft Excel spreadsheets are perhaps the easiest to use for this purpose, because tables are easily presented and most of the necessary statistical functions are built in to the spreadsheets. Staff responsible for generating descriptive statistics should understand pivot-tables, how to use functions to draw graphs, and how to use logical and lookup functions (such as IF, ISBLANK, and VLOOKUP in Excel). **[cf. 5.2.1]**

6.5 The descriptive statistical tasks can be grouped by Master Table (recent victim, mined area or community), as most refer to either mined area or recent victim data, or summaries of these for districts and provinces. Some statistics refer to other types of data or include two types. For example, LIS reporting requires estimates of mine victim rates in a year and in Yemen, statistics were needed on total surface of all mined areas by vegetation and terrain types. The creation of tables from two or more Master Tables does not occur often, and procedures for how to construct these tables are included in the appendix. Data for non-affected communities, also present in IMSMA and the Master Tables, needs to kept separate from affected communities' data **[cf. 5.2.1]**

6.6 In the past, case studies have been used for human interest material and not for highlighting statistical features of mine-affected communities. Statistical profiles that illustrated differences in community background in Yemen were helpful, because they showed correlations between certain variables, such as low accident probability with high community infrastructure. **[cf. 5.3.1]**

7.0 Materials

7.1 IMSMA (required)

7.2 relevant software (required)

8.0 Attachments

[NB: These are documents from different surveys, reports or old protocols that have been attached to this protocol in case they are found useful, interesting or relevant. They are purely optional reading.]

8.1 P16c – Query Train Template

8.2 “Overview and Conceptual Considerations”, by Aldo Benini

8.3 “Information Flow and Conversion”, by Aldo Benini

8.4 Data Analysis Operating Manual, by Aldo Benini

10.0 Related Protocols and Advisories

10.1 P08 -- Impact Scoring

11.0 References

11.1 P09 – Data Analysis is included under objective 3.4,4.1 and output 13 of the generic timeline.

12.0 Revisions

12.1 Written by Aldo Benini 01 01 16

12.2 Edited by Dann Naseemullah 03 01 21

Attachments

1.0 Query Train Template for Data Conversion ('16c')

[In MS Access file attached]

2.0 Overview and Conceptual Considerations

Introduction

The "Global Landmine Survey Initiative" document of October 2000 sets the objectives of the impact survey data analysis. It defines four areas on which country survey reports will elaborate:

Scope of the Problem - The survey will provide absolute and relative numbers of persons impacted by landmines within a particular country. This can be done at national, district, sub-district, and community levels. Similar information will be provided in terms of basic victim profiles, areas of land contamination, and blocked infrastructure.

Impact on Communities - Central to the analysis conducted during the survey will be an examination of community impact, by community type and impact indicator type. Profiles will be developed of low-, medium-, and high-impact communities.

Causality Analysis - Efforts will be made within the survey to explain adaptive behaviours and to determine if any strong relationships exist between certain descriptive aspects of a community and the community's differing likelihood of experiencing impact from mines and UXO.

Implications for Mine Action - In addition to noting which communities have received some form of mine action activity in the past, the survey will collect and present a number of facts critical to subsequent mine action. These will include an analysis of terrain type and vegetation cover, breakdown of contaminated areas by munitions type (UXO or mines), recording of known safe locations for viewing a SHA, and a record of digital photos."

The necessary analyses to fill this program for the Yemen impact survey were done in summer 2000, and the report was published later in the same year. Similar types of analyses have been performed in countries with completed surveys, i.e. Chad, Thailand, Mozambique, and will be performed in other countries with such surveys. Some variation occurred in country specifics. They were elaborated in cooperation between survey implementation organizations, the Survey Action Center, and outside consultants called to answer specialist questions.

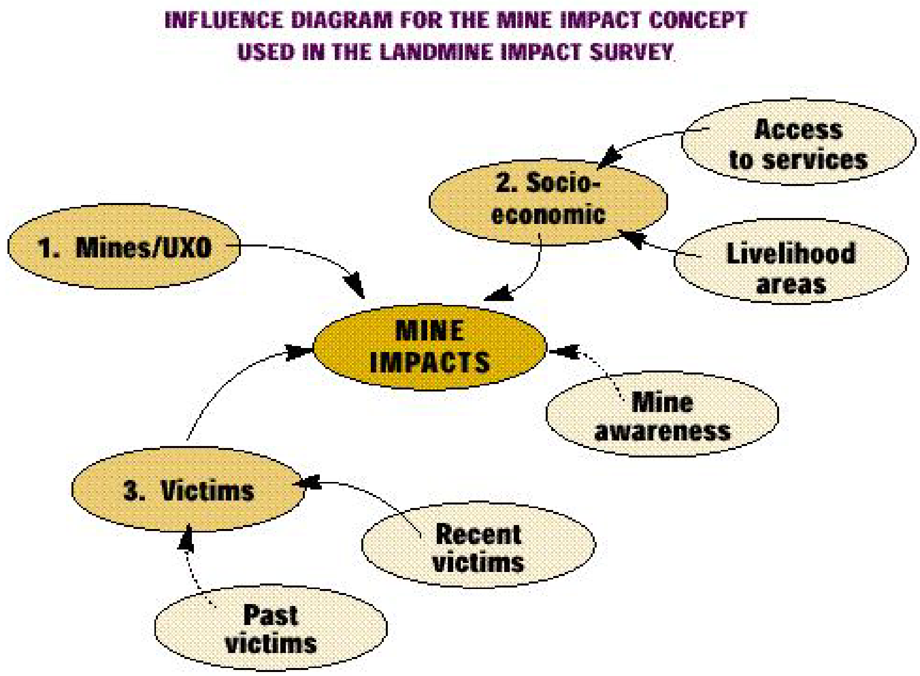
This protocol aims to help accelerate the data analysis phase of the Impact Surveys. In general, most of the analysis types are descriptive, and thus the required skills are modest. However, because of the relational character of the database, the preparation of the master data tables is not a trivial task. It can be greatly abbreviated by producing in advance query templates that can be imported into country IMSMA data pools, and which, with a minimum of adaptation, will then produce those tables. The templates should be imported, tested and adapted in the pilot test phase, and the analyses should be practiced latest on the entered pilot test data pool.

The result of these queries are three master tables, one on affected communities, one on mined areas and one on recent victims, that are indexed in several ways. Each of their records is marked by its unique IMSMA identifier, but also carries the full administrative identification as well as geographic coordinates. It is these master tables that provide most of the input to the applications in which statistical analysis takes place technically.

Again, it is worth pointing out that once the master tables have been prepared, most of the work is descriptive, simple and amenable to the kind of skilled database management or socio-economic personnel found in the survey countries. It will still need close supervision and inspection, particularly for sensible interpretation and the detection of inconsistencies. Analytical statistics, in other words, procedures that involve probability models, remain of subordinate importance and, in general, ought to be farmed out unless qualified personnel are involved.

Conceptual considerations

The Impact Survey is conceptually centred on the composite mine impact index. This is a characteristic of the affected communities. In the impact scoring, information is added on SHAs and recent victims. The survey collects information on other related subjects even though these do not affect the score. Examples include past victim counts and the level of mine awareness. In this influence diagram, they are linked with dotted arrows while the score-effective ones generate unbroken arrows:



The computation of the impact scores, and its inclusion in the survey table, take place inside IMSMA. The SAC queries simply copy these numbers. It is possible, of course, to compute the scores outside IMSMA, and we recommend doing so with the help of one of the spreadsheets exported from the queries, together with the weighting factor table, for the development of alternative scoring scenarios. But, while weights can be changed, the scoring formula itself is locked in by protocol and algorithm.

In contrast to the impact scoring, the analytical treatment of the other material offers greater freedom. This is constrained, however, by the desire to produce impact survey reports for all participating countries that follow a comparable master format. The statistical outputs required to fill such a format was produced, in large part, during the Yemen survey. The analysis program that we recommend for other surveys, therefore, simply follows what was done in Yemen, with some reordering for convenience. In practice, some adaptation will be called for in countries where additional information is collected, or where some standard items seem unreliable to a degree that they should be used with great caveats only. For example, the Thailand survey collects more detailed information on mined area boundaries, but may lose some detail on vegetation and terrain, and will need to restructure queries and tables in accordance with the way it stores the mined area information.

The Yemen report: Table of contents

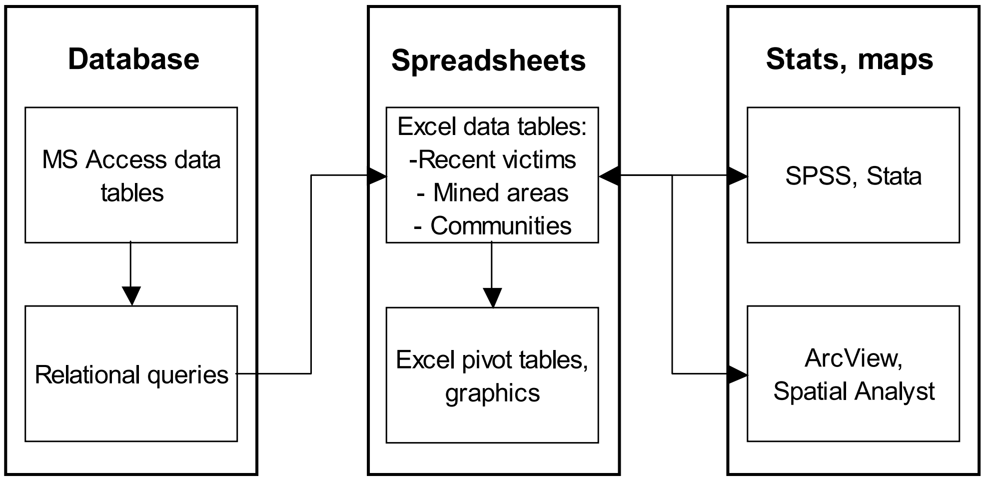
While other country surveys may want to adapt the analytical forms enumerated in this protocol to some degree, the Yemen report furnishes a structure that will be helpful in organizing their material as well:

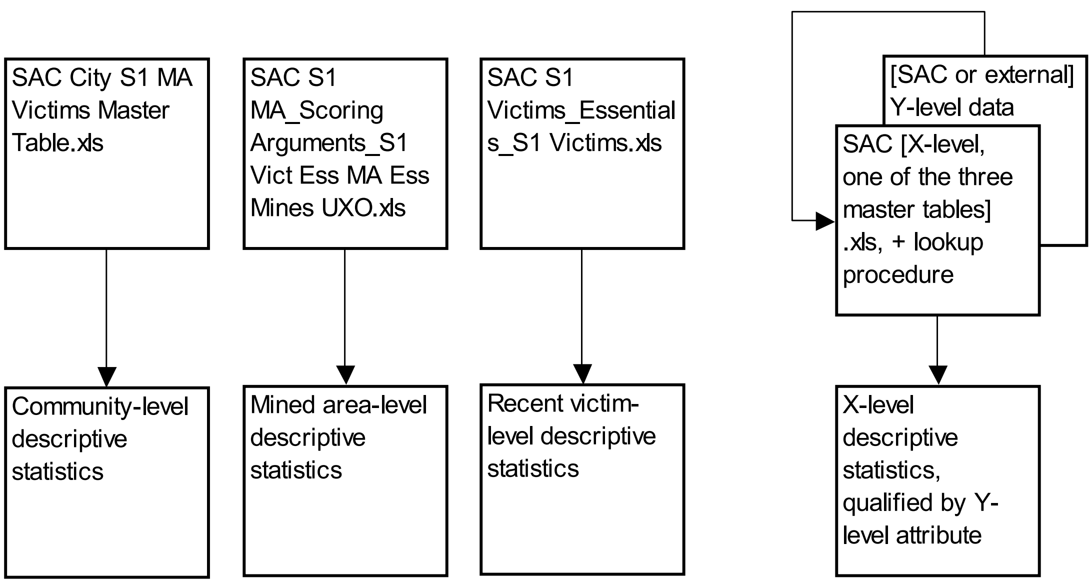
Survey Results and Findings
Scope of the Problem
Impact on Communities
Analysis of Blockage Impacts
Summary of Past Mine Action
Factors Influencing Mine Clearance
Community Background and Mine Effects
Consequences for Mine Action
Background and Methodology
Project Timeline
Key Participants
Administrative Structures
Finances
Yemen Methodology
Attachments and Supporting Documentation
Supporting Analysis
Explanation on Scoring, Weighting and Classifying Communities
Survey Team Leader Report
Case Studies
Estimation of Prevalence of Mine-Affected Communities in Yemen

The development of an analysis program from previous practice, rather than from a unified substantive theory on mine-affected communities, may not satisfy some more theoretically inclined researchers, but the Yemen precedent is considered to have adequately met the agenda of the Global Landmine Survey Initiative, and SAC is more than willing to support efforts that its partners may take in the way of enhancing the analytical substance of country surveys within this agenda.

2.0 Information Flow and conversion

The following figure illustrates the flow of data between applications. Software products are those used in the Yemen data pool analysis; other than for IMSMA, which is written in MS Access and uses ArcView for its GIS extension, SAC does not endorse any of them in particular.





Single-referent forms **Mixed-referent forms**

Surveys may create records for unaffected communities for several reasons. They may import gazetteer data, and/or they may record communities that were visited and turned out to be false positives or true negatives. One has to make sure, when constructing sub-district, district, etc. summaries of affected communities, to count only affected communities. If user-defined fields (and possibly other administrative fields such as visit dates) are used in the IMSMA survey table (tblSurvey1) to characterize the positive/negative status, a filter such as clnPriorityNeutralValue > 0 ("clnPriorityNeutralValue" is the field name for the impact score) has to be applied in queries used to provide the summaries.

3.0 Case Studies / Statistical Profiling

MINE ACCIDENTS AND COMMUNITY BACKGROUND VARIABLES—EXAMPLES

	Communities with highest accident probabilities			Communities with lowest accident probabilities		
	Bait Mashrah, Ibb	Al-Qafleh, Al-Dhale'	Al-Masharih, Al-Dhale'	Tandhoor, Sa'ada	Naseeb Al-Mahjar, Mareb	Imsara, Abyan
Probability of some mine accident in 2 years	0.51	0.50	0.48	0.01	0.01	0.00
Impact score	7	9	12	3	3	2
Actual number of recent victims in the past 2 years	0	1	3	0	0	0
Population	2,000	600	350	150	140	4,000
Access to irrigation blocked	No	Yes	No	No	No	No
Access to drinking water blocked	Yes	Yes	Yes	No	No	No
Year mines last laid	1982	1983	1986	1966	1967	1972
Contaminated area (sq m)	5,000,000	640,000	6,000,000	600	12,000	32,000
Distance to nearest other community with recent victims (km)	3.8	0.5	1.2	21.9	30.3	25.3
Piped water	No	No	No	No	Yes	Yes
Electricity	No	No	No	Yes	Yes	Yes
Telephone	No	No	No	Yes	Yes	No
Health care facility	Yes	No	No	No	Yes	Yes
Primary school	Yes	Yes	Yes	No	Yes	Yes
Secondary school	No	No	No	No	Yes	Yes

While the three communities with the lowest accident probabilities seem trivial situations at first sight, seemingly free of mine impacts, case studies may focus on the history of their development despite the pollution and thus add a depth of perspective that goes beyond the analytical design of the statistical part. How come a small village like Naseeb boasts all of the six selected community facilities?

In other words, statistical profiling can direct attention, in small but potentially enlightening extent, to communities that, by way of case studies, can serve as quasi-experiments in a survey that otherwise misses out on non-affected communities.

4.0 Data Analysis Operating Manual
(Dr. Benini's original appendices)

Queries

Formal principles

Three indexing systems

In each of the data tables of interest, IMSMA provides a unique record identifier, the so-called GUID. This alphanumeric key makes for truly unique identification and for unambiguous links between tables. However, it has no substantive meaning and no link to the real world in which the recorded objects are known by name. The master tables that the SAC queries produce, therefore, come with multiple indexes that allow communicating with precedents (questionnaires, survey staff) and consumers (ArcView, persons who analyse this data and write this report, etc.):

- GUID (truly unique)
- Name and administrative membership
- And, for communities and recent victims, spatial coordinates in decimal degree, which is the only information that ArcView can project

For example, the master query "SAC S1 Victims_Essentials_S1 Victims" comes with no fewer than eleven identification fields:

- Province GUID
- Province
- District GUID
- District
- Sub-District GUID
- Sub-District
- City GUID
- City
- Survey GUID
- Mined Area GUID
- Victim GUID

and would have 13 if the local name of the mined area (often incomplete) and the name of the victim (excluded for privacy reasons) were included.

Aliasing

Calculated and user-defined fields are aliased in ways to clear up ambiguity. Be aware that the upstream queries only show the aliases, not the original field names.

UDF, UDPF and EXT fields

These acronyms stand for the "User-Defined Fields" and "User-Defined Priority Factors" that IMSMA offers as well as external data maintained outside of the IMSMA data structure. The priority factors are fields that can be used for country-specific impact factors; they are linked to the scoring algorithm.

Since their use and meanings will vary from survey to survey, we place them to the rightmost of the query design windows. The same is true in the event some data is maintained outside of the IMSMA data pool (this is the case in Thailand); all external data brought into the queries are placed to the far right of the table. The order of the other fields, in Access and in Excel, will therefore not change when surveys make changes in the definable fields. Any functions in Excel that reference columns of any of the standard fields will therefore not be affected.

In the data tables, the user-defined fields are marked with their substantive alias plus UDF# or UDPF#. External fields can be marked with their substantive alias plus EXT. In queries, only those definable fields are included which are actually used in the tables (undefined fields cannot be included because it is not clear by what operator they would eventually have to be summed up at higher levels).

Dichotomisation using 1 and 0

Check boxes produce the values "–1" for true, "0" for false in the backend binary variables. In accordance with standard social science practice, we converted –1 to +1.

Similarly, the presence of certain classes of munitions is represented as {1; 0}-binaries, with blanks being replaced by "0" (since, in this case, blanks do not stand for "don't know", but for "absent").

Overview

Queries were created at several hierarchical levels in interconnected fashion. They are classified as auxiliary queries and as master queries. Master queries are those that go into statistical procedures after being exported to spreadsheet and statistical applications.

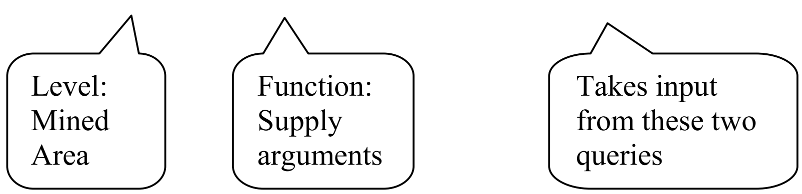
The following queries are supplied with the current version of P 16c ("P16c_QueryReportTemplates_V3_0_020708.mdb"):

- munitions type
- qryDeviceAllSelected
- qryDeviceCustomSelected
- qryDeviceStandardSelected
- qryGIS_SURVEY1_PolarMA
- SAC City S1 MA Victims Master Table
- SAC City_Essentials City
- SAC District Communities MA Surface Victims

- SAC MA Surface by Vegetation and Ground Profile
- SAC Province Communities MA Surface Victims
- SAC S1 City_City Survey Population Comparison
- SAC S1 MA_Essentials_S1 MA S1 PolarMA
- SAC S1 MA_Munitions_Aux
- SAC S1 MA_Scoring Arguments_S1 Vict Ess MA Ess Mines UXO
- SAC S1 MA_Vegetation_Ground_Profile_Index
- SAC S1 Settlement type
- SAC S1 Survey_Essentials S1
- SAC S1 Victims_Essentials_S1 Victims
- SAC Selected device types
- SAC Subdistrict Communities MA Surface Victims
- SAC SubDistricts Vict Aux

The qry* queries were created by IMSMA. The others were created by SAC[1]. The SAC queries were named in a way that the user should recognize the level, the essential function, and some of the contributing queries. For example,

SAC S1 MA_Scoring Arguments_S1 Vict Ess MA Ess Mines UXO



Auxiliary queries

The query "SAC Selected device types" and its predecessors

[1] The query "munitions type" is a special case and may not be necessary any more. See further below.

For completeness, we also list here the modules and reports supplied with 16c:

- SAC Module 1 Vegetation Terrain
- SAC Munitions
- SAC Sub-district Communities MA Surface Victims

The module contents are discussed elsewhere in this protocol.

These replace a group of queries that were described in the old version as follows:

> *The first creates the union of standard IMSMA device types (table "DeviceType")[2] and of user-defined device types (table "tblDeviceTypeCustom"). The second creates a record for each occurrence of a type of munition in a mined area. The third creates a binary for each type of munition occurring anywhere and fills in, for each mined area, the value "1" if this type occurs there, otherwise leaves the cell blank.[3] In other words, the number of records is the number of mined areas for which some type of munition was recorded. In this data pool, this is the case for all the mined areas.*

Some of those were replaced because of the new IMSMA architecture. The upstream queries of "SAC Selected device types" at present are:

- qryDeviceAllSelected
- qryDeviceCustomSelected
- qryDeviceStandardSelected

The internal architecture is difficult to grasp for users other than IMSMA programmers, and we strongly advise not to experiment with their syntax.

16c currently includes also the query

- Munitions type.

This may be superfluous, and the Somaliland test data pool did not require it for the queries to execute. We have not determined whether the query train will run correctly also on any other data pool if this query is eliminated. We therefore retain it for the time being.

The queries "SAC *_Essentials *"

Four queries have this syntax in their names. One each works at the community (called "city" in the IMSMA table lingo), survey, mined area, and recent victim levels respectively. These queries achieve two things:

- They collect all the analytically relevant information from the data tables at the levels for which they are named, leaving out information that is of no concern at all for the time being. For example, most, although not all, variables with an administrative intent are left out. As mentioned above, some of the fields included are aliased, notably the user-defined ones.

[2] This query is a replication of qryAllDeviceTypes found in the frontend of the IMSMA field module. The query will not work on an exported data pool, exports do not include the tables that form the foundation of this query.
[3] This query was supplied by the IMSMA development team.

- They collect the names and GUIDs for all the levels above, starting with the province. This creates tables that make for quick orientation and searches for verification.

For example, the following list displays the first variables of the query "SAC S1 Survey_Essentials S1":

Name	Type	Size	Function
Province GUID	Text	38	Identification
Province	Text	50	Identification
District GUID	Text	38	Identification
District	Text	50	Identification
SubDistrict GUID	Text	38	Identification
SubDistrict	Text	50	Identification
City GUID	Text	38	Identification
City	Text	50	Identification
Survey GUID	Text	38	Identification
Start Date	Date/Time	8	Administrative
End Date	Date/Time	8	Administrative
Last Conflict Year	Integer	2	Analytical
Year Mines First Laid	Integer	2	Analytical
[etc.]			

Of the four "SAC * Essentials*" queries, the one about the recent victims "SAC S1 Victims_Essentials_S1 Victims" is also a master query at the same time because it already holds all the relevant information for victim-level analyses.

The query "SAC S1 MA_Munitions_ Aux"

This is built on the query "SAC Sv1Q1_Crosstab". It creates a record for each mined area with variables for the presence of munitions of a certain nature: Anti-tank mines, anti-personnel mines, UXO, unknown munitions.

These variables are calculated fields using custom functions, which are stored in the SAC Munitions module. For example, the binary variable "Has AP" calls the function HasAP to which this arguments is passed: ([SAC Sv1Q1_Crosstab]![Anti-Personnel][4]. The code for the function:

```
Function HasAP(AntiPersonnel) As Double
If IsNull(AntiPersonnel) = True Then AntiPersonnel = 0
If AntiPersonnel > 0 Then HasAP = 1 Else HasAP = 0
End Function
```

assures that, if any specific types of anti-personnel mines are present (there may have been several recorded for the given mined area!), the presence of

[4] Caution! Do not use reserved symbols or words such as the hyphen when you declare variables in the VBA functions and macros. However, in the expression builder that prompts you to supply the arguments, use exactly the field names as presented, e.g. "Anti-Personnel" with the hyphen.

anti-personnel mines is true. The logic for the other broad types of munitions, i.e. anti-tank mines and UXO, is analogous.

The query "SAC S1 MA_Essentials_S1 MA S1 PolarMA"

This query is centered on the table "tblSurvey1MinedArea". Besides the essential data on the mined areas, it includes the identifications of the higher levels.

Another important function of this query is to recode as 1s or 0s two groups of binary variables:

- the impact variables that supply arguments for the scoring
- vegetation and terrain variables.

The query "SAC S1 MA_Vegetation_Ground_Profile_Index"

This query uses the two functions[5]:

```
Function IndexDifficultyVegetation(Short, Tall, Bush, Tree, Other, None, Unknown As Double) As Double

IndexDifficultyVegetation = None + 2 * Short + 4 * Tall + 8 * Bush + 8 * Tree + 32 * Other + 0 * Unknown

End Function
```

and

```
Function IndexDifficultyGroundProfile(Flat, Hill, Ridge, Gully, Other, Unknown As Double) As Double

IndexDifficultyGroundProfile = Flat + 2 * (Hill + Ridge + Gully) + 8 * Other + 0 * Unknown

End Function
```

to hold these two indexes ready for "SAC S1 MA_Scoring Arguments_S1 Vict Ess MA Ess Mines UXO" so that here the numbers can be translated into meaningful text on the degree of difficulty that vegetation and ground profile present to deminers.

---

[5] Readers have asked why in

IndexDifficultyVegetation = None + 2 * Short + 4 * Tall + 8 * Bush + 8 * Tree + 32 * Other + 0 * Unknown

the coefficients for bush and tree are the same, expecting 16 for "tree". We set both to 8 when deminers explained that bushes and trees were not significantly different regarding terrain difficulty.

The query "SAC S1 City_City Survey Population Comparison"

This is an auxiliary query. It is used for surveys that hold data on the population of the communities from two different sources, say, census data and key informants estimates, and stores it in different tables, "tblCity" and "tblSurvey1". This is the case of Chad.

Province, district, subdistrict summary queries

The four queries

- SAC SubDistricts Vict Aux
- SAC Subdistrict Communities MA Surface Victims
- SAC District Communities MA Surface Victims
- SAC Province Communities MA Surface Victims

create summaries of affected communities and their populations at their respective levels, together with information on the number of mined area, estimated contaminated surface, and four victim variables plus the total victim number. "SAC SubDistricts Vict Aux" is the foundation for the other three; it is needed so that the victims can be totaled in those three.

If there are intermediate tiers between this level and the community, their number is given. These summaries are more convenient to construct as MS Access queries than in Excel. For example, if a pivot table is demanded in Excel to give, say, the contaminated area per province, and the number of affected districts and subdistricts should also be given, then copy these administrative figures from "SAC Provinces_with_surveys City S1 MA Victims Master" into the pivot result.

The report template "SAC Subdistrict Communities MA Surface Victims" is based on the query by the same name.

Master queries

The query "SAC City S1 MA Victims Master Table"

This is one of the final outputs as far as query building in MS Access goes. It supplies, by exporting it to MS Excel, the data table with which most of the community-level analysis is done, either in Excel or in statistical programs that import from Excel.

The query is centered on "SAC S1 Survey_Essentials S1", all of whose variables it includes. From "above", i.e. from the query "SAC City_Essentials City" it takes information on the administrative status of the community, the settlement type as well as from user-defined fields [Some surveys store certain community background variables in "tblCity", in which case this query would hold it too.].

From “below”, it fetches all the scoring arguments as well as some other information on all the mined areas associated with a community, from the query “SAC S1 MA_Scoring Arguments_S1 Vict Ess MA Ess Mines UXO”. It uses different totaling rules in order to combine the values of several mined areas into one for the community, such as count for the number of mined areas, sum for their total estimated area and for the recent victims, minimum for the distance to the nearest area, and maximum for the binary impact variables.

The variables of the query “SAC City S1 MA Victims Master Table”

The spreadsheet table based on this query will arguably support the majority of analyses conducted for the Impact Survey report in any country. The full tableau of field definitions, therefore, is given here:

Name	Type	Size	Function
Province GUID	Text	38	Identification
Province	Text	50	Identification
District GUID	Text	38	Identification
District	Text	50	Identification
SubDistrict GUID	Text	38	Identification
SubDistrict	Text	50	Identification
City GUID	Text	38	Identification
City	Text	50	Identification
Survey GUID	Text	38	Identification
Start Date	Date/Time	8	Administrative
End Date	Date/Time	8	Administrative
Last Conflict Year	Integer	2	Analytical
Year Mines First Laid	Integer	2	Analytical
Year Mines Last Laid	Integer	2	Analytical
Conflict History	Memo	N/A	Analytical
Military Activity	Text	25	Analytical
Year Settlement Established	Text	50	Analytical
Longitude (dd)	Double	8	GIS
Latitude (dd)	Double	8	GIS
Referance Point Discription	Memo	N/A	Administrative
Interview Language	Text	50	Administrative
Recent Mine Awareness Education	Text	25	Analytical
Old Victims Killed	Integer	2	Analytical
Old Victims Injured	Integer	2	Analytical
Recent Victim Assistance	Text	25	Analytical
Current Population	Single	4	Analytical
Pre-War Popullation	Single	4	Analytical

Marking and Survey	Text	25	Analytical
Mine Clearance	Text	25	Analytical
Local Clearance	Text	25	Analytical
Effects of Local Clearance	Text	255	Analytical
Piped Water	Long Integer	4	Analytical
Well or Handpumped Water	Long Integer	4	Analytical
Auto Fuel	Long Integer	4	Analytical
Weekly Market	Long Integer	4	Analytical
Medical Facility	Text	25	Analytical
Primary School	Integer	2	Analytical
Secondary School	Integer	2	Analytical
Collective Organizations	Long Integer	4	Analytical
Collective Organizations Text	Text	255	Analytical
Economic Base	Text	25	Analytical
Other Economic Base	Text	255	Analytical
Survey Comments	Memo	N/A	Analytical
Impact Category	Text	25	Analytical
Impact Score	Single	4	Analytical
Corrected Impact Classification	Text	25	Analytical
Corrected Impact Score	Single	4	Analytical
Impact Correction Factor	Single	4	Analytical
Impact Correction Justification	Memo	N/A	Analytical
Impact Description	Text	255	Analytical
Change of Severity	Text	25	Analytical
Ordinary Village	Long Integer	4	Analytical
City Type	Text	25	Analytical
Mined Areas	Long Integer	4	Analytical
Distance to Nearest Mined Area	Single	4	Analytical
Total Estimated Surface	Double	8	Analytical
Irrigated Crop	Long Integer	4	Analytical
Rainfed Crop	Long Integer	4	Analytical
Fixed Pasture	Long Integer	4	Analytical
Migratory Pasture	Long Integer	4	Analytical
Drinking Water	Long Integer	4	Analytical
Other Water Uses	Long Integer	4	Analytical
Non-Agricultural Land	Long Integer	4	Analytical
Housing	Long Integer	4	Analytical
Roads to Administrative Centers	Long Integer	4	Analytical
Other Roads	Long Integer	4	Analytical
Other Infrastructure	Long Integer	4	Analytical
Total Recent Victims	Double	8	Analytical

Total Recent Fatalities	Double	8	Analytical
Total Recent Injured	Double	8	Analytical
Has AP	Double	8	Analytical
Has AT	Double	8	Analytical
Has UXO	Double	8	Analytical
Has Unknown	Double	8	Analytical
Has Mines	Double	8	Analytical
UDF1	Long Integer	4	
UDF2	Long Integer	4	
UDF3	Text	50	
UDF4	Text	50	
UDF5	Memo	N/A	

The query "SAC S1 MA_Scoring Arguments_S1 Vict Ess MA Ess Mines UXO"

The query has two functions:

- It holds the information needed for almost all analyses required at the mined area level. For this reason, it is ranked as a master table.
- It assembles, at the level of the mined areas, all the information that needs to be passed on subsequently for the scoring at the community level.

The query is centered on its predecessor "SAC S1 MA_Essentials_S1 MA S1 PolarMA". It adds information on the presence of mines and/or UXO and counts in the number of recent victims. It receives also vegetation and ground profile information, but incompletely, reduced to the vegetations and ground profile difficulty indices. It then transforms these numerical expressions into text, using the functions:

```
Function VegetationIndex(Index)
Select Case Index
Case 0
VegetationIndex = "Unknown"
Case 1
VegetationIndex = "None"
Case 2, 3
VegetationIndex = "Short grass only"
Case 4 To 7
VegetationIndex = "Tall grass, at least some"
Case 8 To 31
VegetationIndex = "Bushes or trees, at least some"
Case Else
VegetationIndex = "Other"
End Select
```

```
End Function
```

and

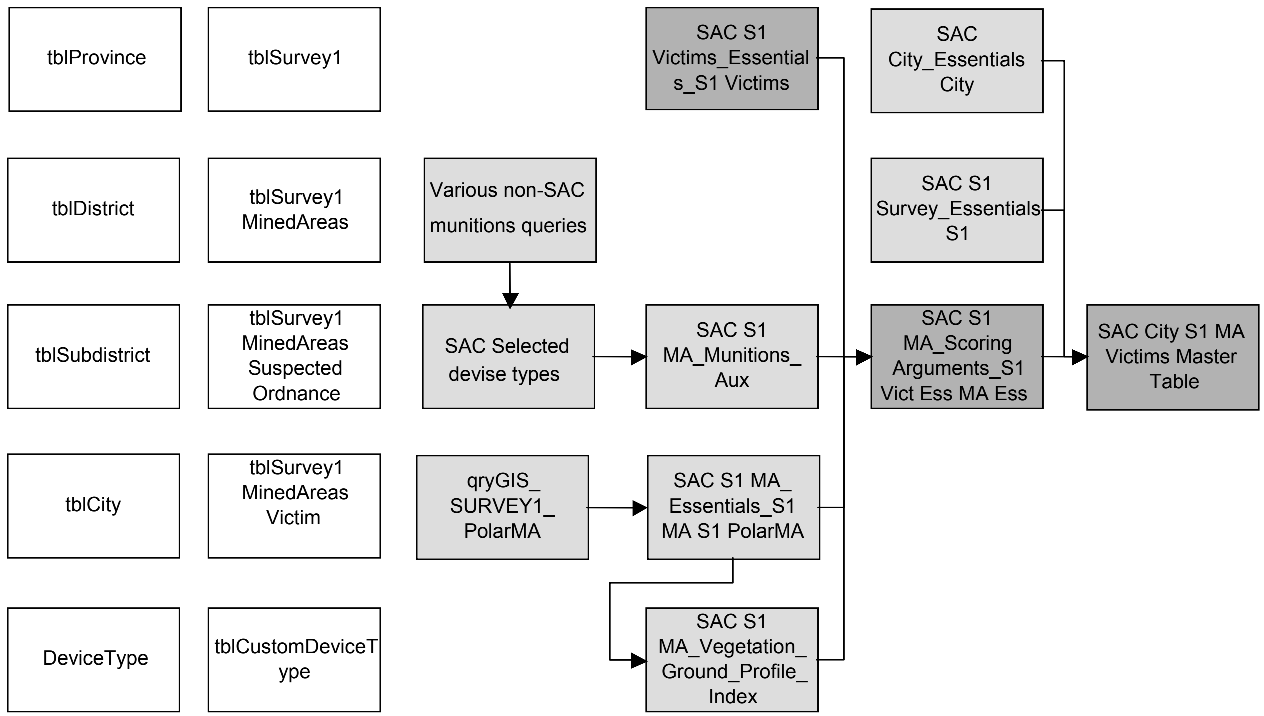
```
Function TerrainIndex(Index)
Select Case Index
Case 0
TerrainIndex = "Unknown"
Case 1
TerrainIndex = "Flat land only"
Case 2 To 7
TerrainIndex = "Contains gullies, hillside, or ridge"
Case Else
TerrainIndex = "Other"
End Select
End Function
```

The query “SAC S1 Victims_Essentials_S1 Victims”

This query holds all the important data on recent victims. It is centered on the table “tblSurvey1MinedAreaVictim”. It is useful because it includes, as the community master query does, the triple indexing (GUID, administrative and geographic coordinates) and recodes as 1s and 0s the values of variables that appear as checkboxes in the IMSMA frontend. The coordinates are those of the community center, not of the mined area.

This table is used for social and accident profiling.

Data tables and queries, hierarchical order



Boxes symbolizing data tables have white background. Auxiliary queries are shaded light grey. Master queries are dark grey. Arrows indicate the hierarchical arrangements among queries. Relationships to the data tables are not shown for space reasons.

Other queries, of minor importance, are not shown here.

How to install the queries

As well as the associated modules and reports:

1. Make a back-up copy of your existing **FieldModuleDataPool**
2. Open the existing **FieldModuleDataPool** by doubleclicking the associated shortcut in your …\IMSMA directory
3. Connect as **AdminIMSMA**
4. In the **File** menu, choose **External Data** then **Import** and select the **P16c_QueryReportTemplates_V3_0_020708.mdb** (in the directory where you are storing it)
   ➔ press the **Queries** tab, select **all** the queries, press **OK**
5. Repeat **step 4** *and*
   ➔ press the **Modules** tab, select **all** the modules, press **OK**
6. Repeat **step 4** *and*
   ➔ press the **Reports** tab, select **all** the reports, press **OK**

Specific analysis forms to cover

Descriptives community-level

Affected communities, with populations

By province

Summary tables showing the number of affected communities, their populations, number and total surface of mined areas, number of victims, and perhaps other variables (e.g. year mines laid latest) are easy to create by way of Excel Pivot tables off "SAC City S1 MA Victims Master Table.xls".

A difficulty arises when the numbers of involved intermediate tier units, e.g. affected districts and subdistricts, are demanded as part of the provincial summary. The Excel Pivot wizard does not offer an operator to count distinct districts and subdistricts only (asking for the count of, e.g., district GUIDs would simply repeat the number of communities). This has to be resolved by forming a train of queries in Access building one upon another, and starting with the first tier above the community (normally the subdistrict) and moving to the province[6]. The following screen shots illustrate part of this process, from subdistrict-level to district-level pivot, building on the community-level master query (SAC City S1 MA Victims Master Table). SAC provides a query "SAC SubDistricts_with_surveys City S1 MA Victims Master"; a segment of the result from a Chad data pool is given here:

[6] The process can be done in Excel as well, but only very tediously so, with a lookup procedure adding the names of the next higher tier units from subdistrict and district dictionary tables that would yet have to be constructed.

Province	District	Subdistrict	Communities	Population affec
B.E.T	BORKOU	COMMUNE DE	3	7400
B.E.T	BORKOU	ANNAKAZA	11	4185
B.E.T	BORKOU	DONZA	15	8765
B.E.T	BORKOU	KAMADJA	4	7000
B.E.T	BORKOU	TCHOMOGO	1	260
CHARI-BAGUIRMI	MOITO	MOITO	3	4900
CHARI-BAGUIRMI	MASSAKORY	KANEMBOU	1	250
CHARI-BAGUIRMI	MASSAKORY	HADDAD	3	500

The following query ("SAC Districts_with_surveys City S1 MA Victims Master") summarizes at the district level:

Province	District	Subdistricts	Communities	Population
B.E.T	BORKOU	5	34	27610
CHARI-BAGUIRMI	MASSAKORY	2	4	750
CHARI-BAGUIRMI	MOITO	1	3	4900

The district table is the input to the province-level query "SAC Provinces_with_surveys City S1 MA Victims Master" (not shown here).

It would be feasible to use these summaries as labeling devices for Excel pivots if other summary-type analyses at these levels were required. Export them, copy them as sheets into the workbook in which the pivot tables are created, reorder the columns so that the subdistricts are followed by districts and then provinces, name the tables as ranges, and use them as lookup tables in the VLOOKUP procedures in order to fetch administrative names.

By settlement type

A cross-tabulation of this type will be useful where settlement type information is complete and reliable:

Settlement type	Affected communities	Population	Mean population
Urban	10	93,640	9,364
Suburban	24	113,960	4,748
Compact village	230	323,216	1,405
Dispersed village	283	284,753	1,006
Seasonal village	8	737	92
Nomadic	32	10,932	342
Other	5	556	111
Total	**592**	**827,794**	**1,398**

The example is from Yemen. The table is easy to construct as a pivot table off "SAC City S1 MA Victims Master Table.xls". Rows have to be reordered to bring out the continuum from urban to nomadic, and the residual category "Other" should be illustrated in a footnote, e.g.: "E.g., an ocean fishermen's staging post, a former police station, a bus stop with stores, most of them in Aden and Abyan near or on the coast".

By population size, histogram

The population of affected communities is expected to be lognormally distributed. In "SAC City S1 MA Victims Master Table.xls", one is to create this variable, using the function LOG10 (some other variables, such as the mined area surface and the distance to the nearest community with recent victims, follow the same distribution and should also be transformed). The presentation should be by way of a histogram. This can be done in Excel as a bar graph, but more conveniently in SPSS, which also offers a normal curve overlay. Key statistics, such as median, minimum, and maximum population size should be given in untransformed values. SPSS users will also want to run a box plot to visualize outliers. Also, inspect the lower end of the histogram for possible truncation of small communities, which, if found, may be explained by survey bias, grouping of small communities, or administrative regulations for minimum community size.

By impact score, by impact category

While the numbers of affected communities by impact score and, based on that, impact category arguably are among the most important outputs of the Survey, constructing the score histogram and the category summary is trivial. One note of caution, though: In Excel, when using the function "FREQUENCY", recall that the ranges are defined by their upper limit; read the Help text. In SPSS, frequency tables do no give categories with no occurrences; if you want a complete table also with the zero-occurrence score values, then you have to export the table and insert rows for those values in Excel.

Occasionally, when listings of communities are demanded contingent on some property, it is efficient to create them in SPSS.

By types of areas blocked, incl. no. mined areas, surface

The table "SAC City S1 MA Victims Master Table.xls" contains binaries for each type of blockage. Summaries including the number of mined areas and their combined surface can be done using pivot tables. However, for every type of blockage variable, a separate pivot table has to be constructed. This can be done most efficiently by using only one temporary table, changing the row field in the Wizard till all types are done, and collecting the results in a separate area or sheet with the help of "Copy – paste special, values only".

By exposure to past mine action

The table "SAC City S1 MA Victims Master Table.xls" contains binaries for each type of past mine action. Unless summaries with other variables or cross-tabulations are demanded (say, by province), simply use the summing function, otherwise pivot tables. In keeping with the rule that an Excel data sheet should not contain calculated rows, place those sum formulae in a different sheet or remove them after copying the results.

By nature of munitions

It is important to be clear whether statements on the occurrence of types of munitions are made at the mined area or the community level. When the community level is meant, the table “SAC City S1 MA Victims Master Table.xls” contains binaries for the occurrence, anywhere in the community, of anti-personnel mines, anti-tank mines and UXO. The same pivot table can handle all their combinations (drag all the munitions binaries into the row area of the pivot wizard), the number of communities, and the total mined area surface in the communities belonging to each of the various munitions combinations. By using the grouping technique, a combined category for all mixed system pollution form can be created if desired.

That situation is different from mixed form analyses. These are rare. They occur when some mined area level statistic is demanded, qualified by values of a variable defined at another level – usually the community. As mentioned elsewhere, in such cases the qualifying property has to be linked to the mined area records using special techniques. The technique of choice in Excel is the function VLOOKUP. Since the survey GUID is part of the “SAC S1 MA_Scoring Arguments_S1 Vict Ess MA Ess Mines UXO.xls” master table, it is the ideal look-up variable for the function. Care has to be taken to name a look-up range within “SAC City S1 MA Victims Master Table.xls” with the survey GUID being the leftmost column.

By age of conflict, histogram

We use the variable “clnMinesLastPlanted” in the IMSMA survey table to characterize the age of the conflict as the time the community has had in order to adapt to the mine pollution. This variable, not the alternative “clnLastConflict”, is used in the SAC queries and is aliased as “Year Mines Last Laid”. The number of communities who saw mines last laid in year x can best be represented as a histogram.

In multi-variate models of the Yemen analysis, the three conflicts were represented simply by the number of years that passed since the last mine laying. This may not be appropriate for all situations. An alternative is to form binary variables for each major conflict (a community may have been subjected to violence in several conflicts, but the information may be incomplete), and characterize the community by the one conflict or conflict period in which mines were last laid. If there were n conflicts, n binaries may be created, but only (n – 1) can be included in a regression model, in order to avoid linear dependency.

By institutional endowment

The survey collects a small amount of background information on the affected communities. Except for population, this information is not used in stand-along fashion, but is valuable as potential causal factors for the measure and kind of adaptation that communities achieve to the mine pollution. We attempt to further reduce the complexity of this background information through scale formation or factor analysis, and in the summary table for these analyses (e.g. factor loadings) the frequencies of institutional traits should be displayed. They are simple to calculate by summing the "1"s in each of the binaries and dividing by the number of communities. As an example, the frequencies of eight such characteristics in Yemen are shown in the factor analysis summary on page 43.

By distance to the nearest other community with recent victims

This measure is a proxy for the spatial clustering of high-impact communities. It was a significant determinant for communities in Yemen to have had recent mine accidents, and may operate similarly in other contexts. It was the only explicitly spatial variable in the causality model used in the Yemen analysis. Its validity is derived from the fact that the local mining intensity does not respect community boundaries, but historically followed the logic and dynamics of the local conflicts. Moreover, mined areas are often visited by members of neighboring communities.

The variable can be calculated in Excel using a macro. At first, an auxiliary table is created to hold this information:

- City GUID
- City
- clnLongitude
- clnLatitude
- Recent victims

The macro then returns these variables:

- GUIDNearestCityWithRecentVictims
- Distance
- NoCommunitiesWithRecentVictimsWithin10kmRadius

The macro as used for Yemen is appended (see page 29). It is easy to adapt for other countries.

The validity of this measure may vary from country to country. The impact surveys are one-country surveys, and the distance is defined to the nearest other *domestic* community with some recent victims. In countries with heavily mined border communities and considerable cross-border traffic, however, vicinity to another mine-affected community across the border may have similar or greater significance than the neighborhood to affected communities within the country. This consideration may apply to countries like Thailand and Cambodia.

By district, with population, mined areas, surface, number of victims (by recent/earlier and fatal/injured)

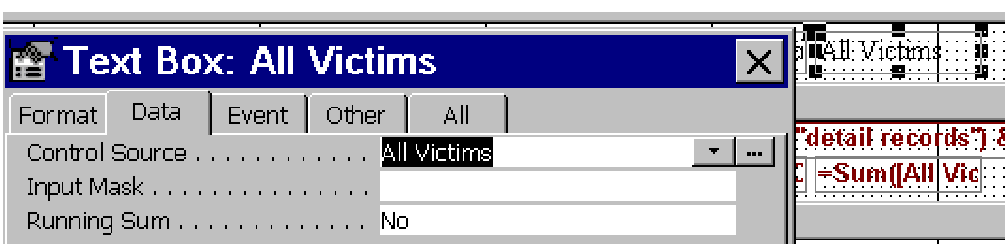
This type of summary report is best illustrated with the following segment taken from the Yemen report:

SUMMARY: COMMUNITIES AND POPULATIONS AFFECTED, BY GOVERNORATE AND DISTRICT

	District	Communities affected	Population affected	Mined areas	Contaminated surface area (sq m)	Victims: Recently killed	Recently injured	Killed earlier	Injured earlier	All victims
Abyan	Khanfar	15	7,027	27	83,988,200	6	4	57	37	104
	Lowdar	3	4,525	4	441,500	2	0	17	1	20
	Modya	1	20,000	3	45	0	0	2	1	3
	TOTAL	19	31,552	34	84,429,745	8	4	76	39	127
Aden	Al-Buraiqa	15	43,690	27	54,490,000	3	1	32	45	81
	Al-Mansoora	1	2,000	1	420,000	0	0	6	0	6
	Dar Sa'ad	2	2,900	3	1,610,000	0	0	8	0	8
	Khour Maksar	1	150	1	3,000,000	0	0	0	0	0
	Sheikh Othman	1	950	2	2,410,000	0	0	0	0	0
	TOTAL	20	49,690	34	61,930,000	3	1	46	45	95

This type of report is best created in MS Access. While the Report Wizard makes it a sheer fun trip to create a report off a query, the queries themselves may have to be adapted, particularly in order to include calculated fields correctly. In this case, the report could not be built directly off the subdistrict and then district summary templates that we include with the template base, but had to be replicated onto itself, by first creating a query with everything except "All victims". Then one based on this one, plus "All Victims" calculated as the sum of four victim variables, was built to inform the report. We include both the query and the report, detailed to the subdistrict level, as "SAC Subdistrict Communities MA Surface Victims" in the template collection.

Access allows users to format reports to a great detail, either manually or by right-clicking on the controls and changing their properties. Here is a screenshot example of of a control selected, "All Victims", with the properties dialogue box brought up:

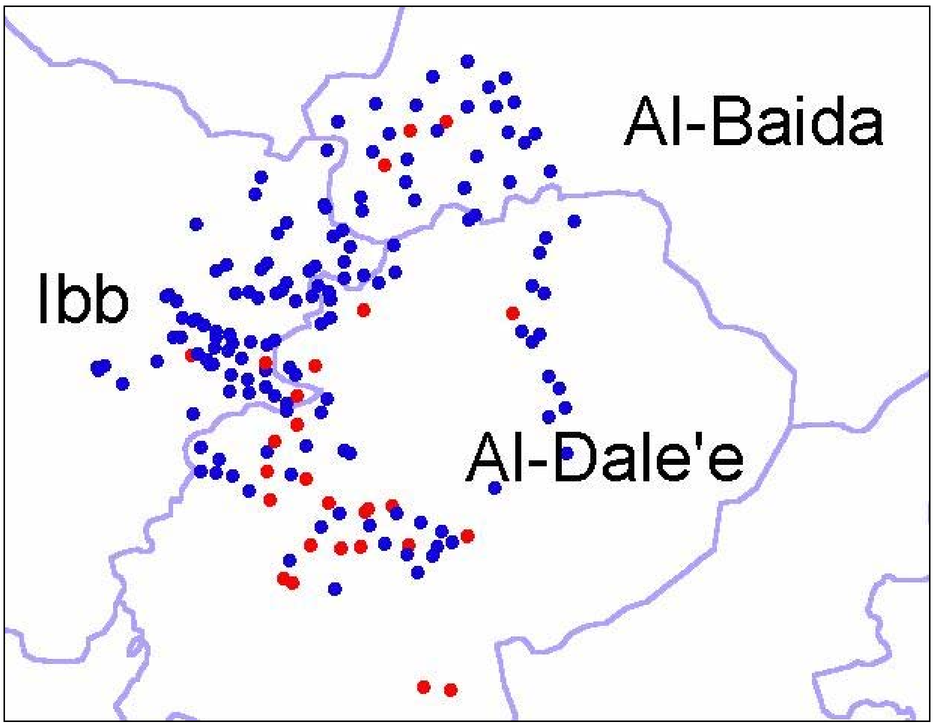


Geographic distribution

In addition to the tabular and histogram outputs described earlier, maps are needed to visualize the spatial clustering and distribution, across the country, of key variables such as

- Number of affected communities (or percent affected population) by district
- Communities by age of conflict
- Communities having some recent victims vs. others with none

For the latter type of map, a binary variable based on the number of recent victims is needed, using the IF-function in Excel. The following picture shows the clustering of red dots (communities with recent victims) for part of the central highland in Yemen. Depending on research interests, other variables could also be displayed for their spatial distribution, such as past mine action.



Macro to compute distance to the nearest other community with recent victims

Create the auxiliary spreadsheet described on page 27, a segment of which is shown here:

	1	2	3	4	5
1	City GUID	City	clnLongitude	clnLatitude	Recent victims
2	{0118AAF2-EC10-11D	Al Sawda	46.77017	14.55122	0
3	{0118AAF9-EC10-11D	Al Dhahair	46.62403	14.64964	0
4	{0118AAFF-EC10-11D	Hubaidh Al Asfal	47.27628	15.32231	0

and name it “DistanceData”. In the same workbook, create another sheet called “Transfer”. Paste into this sheet an exact copy of this following table, in formula view (Tools – Options – View – Window Options: Formulas: checked), when cell R1C1 (or “A1”) is the active cell (in other words, “Great circle method” goes into A1):

Great circle method			
RadiansLatitFrom		Sin(RadiansLatitFrom)	=SIN(RC[-2])
RadiansLatitTo		Sin(RadiansLatitTo)	=SIN(RC[-2])
		Cos(RadiansLatitFrom)	=COS(R[-2]C[-2])
		Cos(RadiansLatitTo)	=COS(R[-2]C[-2])
RadiansLongitDiff		Cos(RadiansLongitDiff)	=COS(RC[-2])
		Distance (radians)	=ACOS(R[-5]C*R[-4]C+R[-3]C*R[-2]C*R[-1]C)
		Distance (degrees)	=(R[-1]C/PI())*180
		Distance (km)	=1.852*60*R[-1]C
			1 minute of arc is 1 nautical mile
			1 nautical mile is 1.852 km

The precision in copying and pasting is needed so that this formula from spherical trigonometry can correctly interact with the following macro. Copy this long piece of code into the VBA editor in a module that you open for this workbook. Read it in the editor and adapt some of its parameters, e.g. number of communities in your DistanceData worksheet, and save the module. Make sure "DistanceData" is the active sheet when you run the macro (press Alt + F8).

```
Sub DistanceCRV()

'Calculates distance from a given community to nearest community
'that has some recent victims, other than itself. Also calculates
'the number of other communities, within a given distance, that
'had some recent victims. Retains the CityGUID of the nearest such
'community.

'This distance is a measure of spatial clustering of landmine
'impacts. In the Yemen zero-inflated Poisson regression model of
'factors influencing the number of recent mine victims, this one
'was significantly related to communities' ability to avoid accidents
'(meaning, have zero recent victims).

'Written by Aldo Benini May 2000
'Modified January 2001, with help from Charles Conley

'Uses the great-circle method with a formula
'taken from http://www.auslig.gov.au/geodesy/datums/distance.htm

'Definitions:
'* South latitudes are negative
'* East longitudes are positive
'* Great circle distance is the shortest distance
'*between two points on a sphere. This coincides with the
'*circumference of a circle which passes through both points
'*and the centre of the sphere.


'Initialize values:

'Number of cases
NoCommunities = 592 'The number of affected communities in Yemen.
```

```
'Radius of circle within which to count other
'communities that had some recent victims
'In kilometers.
Radius = 10

Pi = Application.WorksheetFunction.Pi()

'Outer counter
For Counter1 = 1 To NoCommunities

'Reinitialize distance
'This is needed because the macro asks for distance
'to be modified only if the value at subsequent counter
'stops are smaller.
Distance = 1000000

'Reinitialize number of communities with recent
'victims within circle
WithVictimsInCircle = 0

'Read coordinates of community FROM which to measure
LongitFrom = Cells(Counter1 + 1, 3).Value
LatitFrom = Cells(Counter1 + 1, 4).Value

'Inner counter
For Counter2 = 1 To NoCommunities

'Exclude itself:
If Counter2 <> Counter1 Then

'Read coordinates of community TO which to measure
LongitTo = Cells(Counter2 + 1, 3).Value
LatitTo = Cells(Counter2 + 1, 4).Value

'Determine whether community had any recent victims
HadVictims = Cells(Counter2 + 1, 5).Value

  'If so, calculate distance to this town only for this case:
  If HadVictims > 0 Then

  'At first coordinates need transformation into radians:
  RadiansLongitFrom = LongitFrom * Pi / 180

  RadiansLatitFrom = LatitFrom * Pi / 180
  'This value needs to be written to an auxiliary worksheet
  'called "Transfer", within the same workbook.
  'This contrivance is needed because VBA cannot handle the
  'function "Acos" without calling on a worksheet.
  'When you run the macro, make sure to have the sheet
  '"DistanceData" active, not "Transfer".
  Worksheets("Transfer").Cells(3, 2).Value = RadiansLatitFrom

  RadiansLongitTo = LongitTo * Pi / 180

  RadiansLatitTo = LatitTo * Pi / 180
  'This value too needs to be written to the transfer sheet
  Worksheets("Transfer").Cells(4, 2).Value = RadiansLatitTo

  RadiansLongitDiff = RadiansLongitTo - RadiansLongitFrom
  'This one, too, goes to "Transfer"
```

```
    Worksheets("Transfer").Cells(7, 2).Value = RadiansLongitDiff

    'Now the reverse happens: VBA collects the formula result
    'from the transfer sheet
    '[Note: The coefficients in cell R10C4 are used because
    '1 minute of arc is 1 nautical mile
    '1 nautical mile is 1.852 km.]

    'TempDistance is in kilometers:
    TempDistance = Worksheets("Transfer").Cells(10, 4).Value


        'Calculate number of communities in circle
        If TempDistance <= Radius Then
        'Add 1 to the number of communities with recent
        'victims if this one is within a desired radius
        'from the community of origin:
        WithVictimsInCircle = WithVictimsInCircle + 1
        Else
        End If

        'Decrease distance if value found smaller than previous
        If TempDistance < Distance Then
        Distance = TempDistance
        'Read GUID
        GUIDTo = Cells(Counter2 + 1, 1).FormulaR1C1
        Else
        End If
    'If no recent victims, do not calculate anything
    Else
    End If
Else
End If
Next Counter2


'Pass distance to worksheet cell
Cells(Counter1 + 1, 7).Value = Distance
'Pass city GUID
Cells(Counter1 + 1, 6).Value = GUIDTo
'Pass number of communities with recent victims
'that lie inside circle
Cells(Counter1 + 1, 8).Value = WithVictimsInCircle
Next Counter1
End Sub
```

When you are done running the macro (this will take several minutes depending on the size of the file), save the workbook. Transfer the results back to your community-level master table (if you copy and paste, make sure both tables are sorted identically by CityGUID); import into the required statistical program.

Mined area properties

By number of mined areas

The number of mined areas in a community is given in the table “SAC City S1 MA Victims Master Table.xls”. Summarizing the number of communities by their number of distinct mined areas is, therefore, simple to do by way of a pivot table in Excel or a frequency table in SPSS.

By mined area surface, by vegetation and terrain, filtered to high-impact communities

A mixed form analysis, used in Yemen. The table “SAC S1 MA_Scoring Arguments_S1 Vict Ess MA Ess Mines UXO.xls” includes two variables on the degree of difficulty in vegetation and ground profile. These are based on indexing functions written in the SAC-provided queries. For each community in the high-impact set, a cross-tabulation is produced as shown in the following segment. Note that a given mined area in a community is found only in one of the cells of the cross-tabulation, but any non-blank cell can hold the surfaces added from more than one mined area if several of these combine the same traits.

DETAILS OF MINED AREA TERRAIN AND VEGETATION FOR HIGH-IMPACT COMMUNITIES

	Vegetation	Flat land only	Hillside, ridge and gully	Other	Total
Governorate: ABYAN District: KHANFAR Community: OWAYDAYN	None				
	Short grass only			6,000,000	6,000,000
	Tall grass, at least some				
	Bushes or trees, at least some			6,000,000	6,000,000
	Others				
Governorate: ADEN District: AL-BURAIQA Community: AL-HESWA	None				
	Short grass only				
	Tall grass, at least some				
	Bushes or trees, at least some	160,000			160,000
	Others				
Governorate: AL-BAIDHA District: AL-ZAHER Community: LAJRADI AND AL-MIHSIN VILLAGES	None				
	Short grass only		132,500		132,500
	Tall grass, at least some				
	Bushes or trees, at least some		135,000		135,000
	Others				

We present this only as an example for the need to be clear about the proper unit of analysis – formally, for each community 20 properties are calculated (a 5 x 4 matrix) from information taken from the mined area master table. The actual calculation in Excel is tedious – lifting the community names into the mined area master table, then a repetition of pivot tables with the community name in the page field – and there is no need to do this here. Maybe one or the other of the readers can find a more elegant solution in MS Access – but only if the need arises!

Victim properties

The table

MINE VICTIM SURVEY

Period	Communities involved	Victims		
		Killed	Injured	All
Recent victims	78	57	121	178
Victims of less recent date	474	2,503	2,223	4,726
All victims*	488	2,560	2,344	4,904
Had no victims	104	–	–	–

*The set of communities with some victims, regardless of the date of accidents, is the union of the two period sets, not a simple addition. Some communities had victims in both periods. The victims, however, are mutually exclusive; their numbers add up.

does speak of victims, but has the community as unit of analysis. In addition to summing the numbers of victims, counts are needed on the following variables

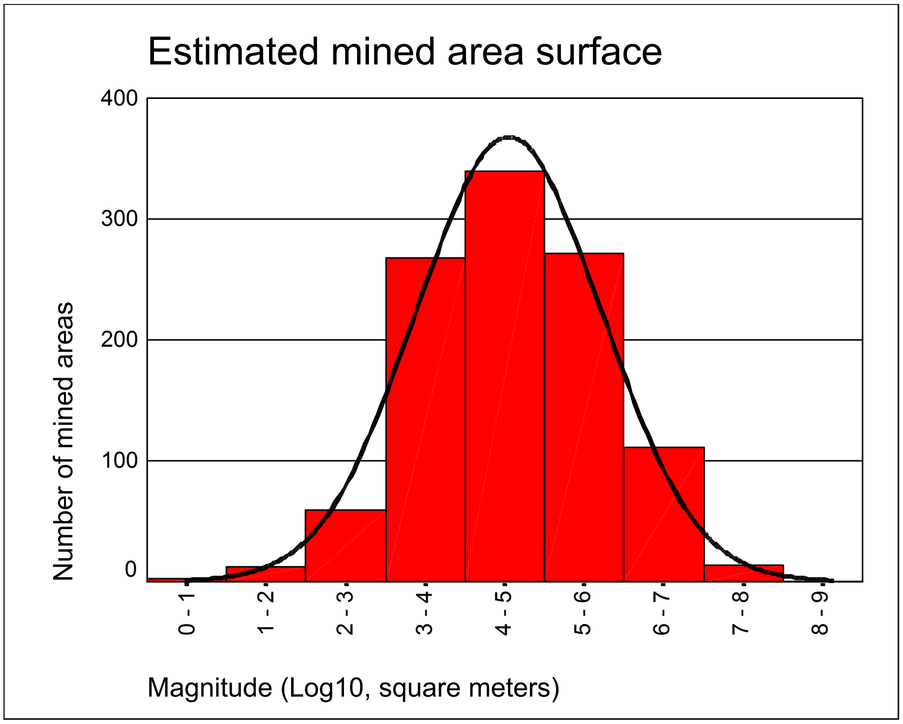
- Had some recent victims.
- Had some earlier victims
- Had some victims of any kind
- Had no victims

These are created in "SAC City S1 MA Victims Master Table.xls" with the IF-fucntion.

Descriptives mined area-level

By surface

Surface estimates span several magnitudes (in Yemen, from 4 square meters to 81 square kilometers). They are lognormally distributed, and a logarithmic transformation should be made in Excel, with descriptives, box plot and histogram produced in SPSS if possible. The histogram from the Yemen analysis is shown as an example:



One of the key statistics that the Impact Survey is expected to supply is the total contaminated area in the country. This has been calculated as the sum of the surface estimates that key informants in the affected communities offered for each and every mined area. However, many a mined area affects several communities, and there is considerable overlap among areas claimed by different communities. In Yemen, this overlap was not taken into account. In future survey analyses, an approximate idea of the overlap factor should be formed by unioning, in ArcView, all mined areas extracted from the relevant IMSMA layer. Then the number of mined areas and their surfaces should be recalculated, together with the reduced total surface. This reduction factor will be very inaccurate, but the inaccuracy from poor projection based on key informant mapping and one-point visual inspections will likely not be greater than the margin of error implicit in the size estimates.

In Thailand, a different approach is being tried out, with mined area polygons that are stored separately from the impact survey part of the IMSMA field module, yet are linked to mined areas enumerated by key informants.

By all combinations of AT, AP, UXO, with no. mined areas, surface, filtered to communities with water problems and/or high mine impact

This is an example of a mixed form. It was used in the Yemen report. Mixed forms arise when the statistic of the basic unit in point is qualified by some property that belongs to a unit of a different level. In this case, the basic unit is the mined area; the community is the qualifier. A mined area is included here if the community to which it belongs has some water source blocked and/or is placed in the high-impact category. In order to calculate this set, the union of these two conditional community sets (water and/or high impact) has to be created by an appropriate IF-statement in the community master table (like =IF([water] + [high]>0,1,0)). Then the VLOOKUP procedure is applied as described elsewhere, to lift the filter variable into the mined area master table. The filtering condition is then used in the page area of the pivot table wizard. The result, for Yemen, was like this:

COMMUNITIES WITH HIGH IMPACT AND WATER RESOURCE BLOCKED— MINED AREA DISTRIBUTION BY CLASS OF MUNITIONS

Type of mine	Mined areas	Contaminated surfaces (sq m)
AP	78	53,668,869
AP,AT	15	209,752,000
AP,AT,UXO	9	35,300,500
AP,UXO	19	1,375,445
AT	4	211,000
AT,UXO	3	2,900,000
UXO	6	2,334,018
TOTAL	**134**	**305,541,832**

As another example of a mixed form, the Yemen report carries a bubble diagram for the numbers and total surface of mined areas for each combination of vegetation and terrain difficulty for the 14 communities classified as high-impact (not shown here).

By vegetation and ground profile, summed for surface

The two variables that the SAC-provided queries compute, at the mined area level, on the degree of difficulty in vegetation and ground profile have been mentioned above. Using those, this cross-tabulation is easy to produce by way of an Excel pivot table. If the residual categories "Other" produce significant surfaces, they should be exemplified in the text.

	Ground profile			
Vegetation	Flat land only	Contains gullies, hillside, or ridge	Other (e.g. moving sand dunes)	Total
None	21.8	6.0	0.1	27.9
Short grass only	292.0	162.5	34.2	488.7
Tall grass, at least some	100.9	3.3	0.0	104.3
Bushes or trees, at least some	119.1	87.5	94.2	300.8
Other	1.0	0.0	0.0	1.0
TOTAL	**534.8**	**259.3**	**128.5**	**922.7**

Other mined area descriptives

Several other statistics were used in the Yemen report. The need to include them may recur for some in other countries, but there may just as likely be a need for different combinations. Therefore, they are mentioned here only summarily:

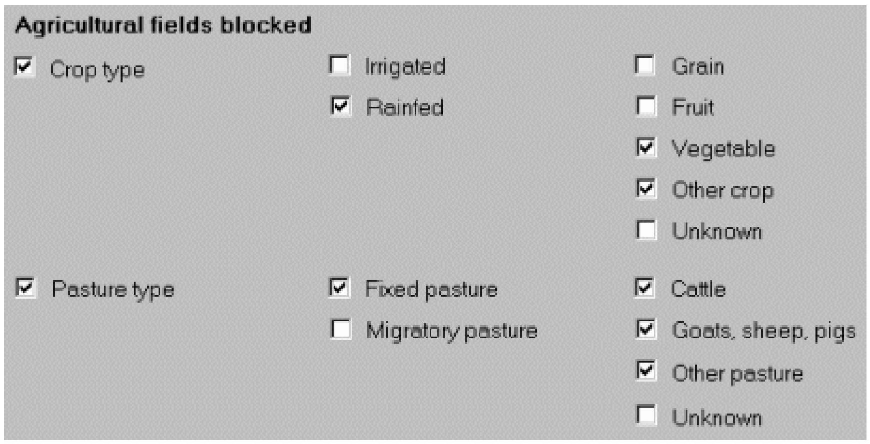
- By surface brackets, and munitions types
- Same, and vegetation difficulty
- By all combinations of AP, AT, UXO, with contaminated surface
- By age (conflict eras), with surface
- Same, filtering out one of the Governorates
- Same, filtered to high-impact communities

They can be produced by way of pivot tables in “SAC S1 MA_Scoring Arguments_S1 Vict Ess MA Ess Mines UXO.xls”, or following the mixed-form procedure. Surface brackets can be created in the Pivot table, with the grouping feature. Filters need to be created in the community-level master table, then brought into the mined area table with VLOOKUP.

Agricultural typologies: Modifications should be done for good reasons only

Some country surveys have expressed dissatisfaction with the agricultural typology used in the survey and in the database. Chad, for example, felt that the date crop, which is important in the north, should receive special mention among the uses of agricultural land that mines or UXO block, and should not be coded only as "fruit" among others.

IMSMA has a three-level structure for the typology. In order to contribute to the impact score, information has to be checked only the first and second levels – the third is optional:



Therefore, if the survey management feels that the typology should include more specific variables, user-defined fields (e.g. "dates", in which also "crop" at the first level, and one of the second-level boxes would have to be checked) or even a user-defined priority factor with its own weight (e.g., "river fisheries", in which case no boxes at the first and second level would need to be checked) could be activated.

While that is technically feasible, one of the costs is reduced compatibility between the IMSMA Field and Global Module, and the loss of some information in transfer from country to UN headquarters. Some of the queries may have to be modified, too. Maintaining the modified structure during IMSMA version upgrades may be problematic.

Descriptives recent victim-level

In the Yemen report, all statistics at this level were produced off the equivalent of "SAC S1 Victims_Essentials_S1 Victims.xls", using pivot tables:

- By age, bar graph
- By age and gender
- By fatality and gender, incl. fatality rate

- By sex, military/civilian status, occupation before
- By activity at time of accident
- By gender and currant occupation of survivors

- By care of survivors, and gender
- By care received, and governorate
- By type of injury, and governorate

Note that in IMSMA, "care received" and "injuries", in order to allow for multiple choices, are recorded as sets of binary variables (check boxes).

Therefore, cross-tabulations (of gender, or geographical unit, or any other variable) with either of those have to be constructed using several pivot tables, or using the same pivot table for temporary results transferred, for the positive value of each binary variable, to a collection area.

Descriptives using other referents

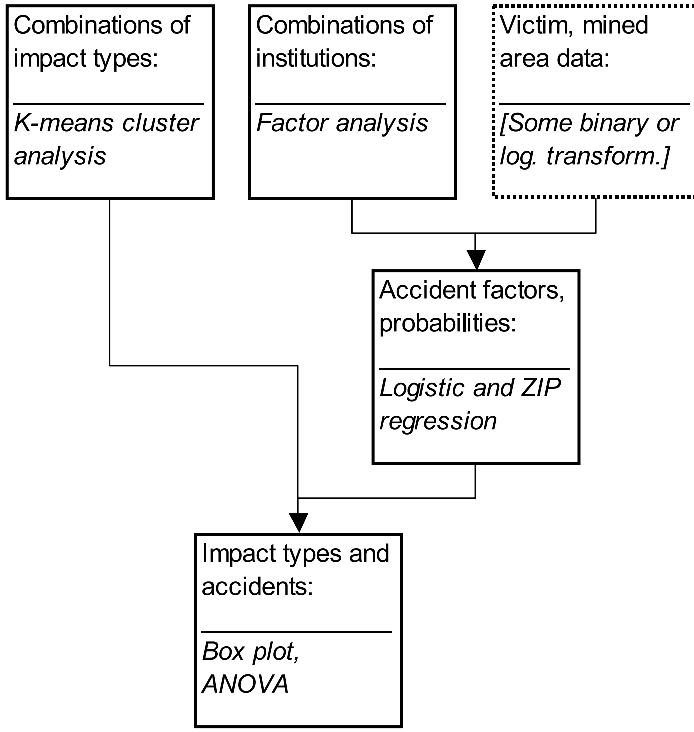
There is only one instance - victim rates based on population other than that of the affected communities. Meaningful rates can be formed for the population of affected districts as well as for the national population[7]. This presupposes that population figures are available for entire districts and for the country. The challenge is not primarily analytical (although there may be inconsistencies between local population estimates such as by the key informants from affected communities and the census). It is administrative in the sense that the survey-implementing organization must start searching for this information early on – not only when the fieldwork is completed.

Uses of analytical statistics

Train of procedures used

The following assumes that the country survey has personnel who are familiar with the procedures invoked, or have access to outside experts who can do them. Some procedures will build on output from others. In the Yemen analysis, the dependencies were as follows:

[7] Strictly speaking, the ratio of mine victims to the population of all affected communities too is a property of the country – there is only one set of affected communities in the country, and countries can, in theory at least, be compared on that rate. What is different is that district population-based and nationwide rates require information on population elements other than those of the surveyed communities.



The following sections annotate each of those.

Combination of impacts, k-means cluster analysis

Communities are different by the kinds of impacts that they combine. In the Yemen analysis, k-means cluster analysis was employed to identify predominant combinations of types of blockage.

This table summarizes the result:

Type	A	B	C	D	E	F	Frequency
Pasture							89%
Rain-fed farms							25%
Local roads and trails							22%
Wood foraging (non-agricult. land)							17%
Water, other purposes							9%
Drinking water							8%
Irrigated farms							6%
Roads to administrative centers							5%
Housing							4%
Communities concerned	*376*	*102*	*27*	*23*	*21*	*43*	

Cells in dark gray designate impacts that were present in the communities of the particular type with a frequency of at least 80 percent. Light gray stands for impacts that were associated with the type much above the overall frequency.

The procedure is not purely descriptive because not all communities possess exactly the properties of the type to which they are assigned. An iterative procedure is used that, for a predefined number of clusters (= types), reevaluates different sets of possible clusters so as to minimize Euclidian distances in the multi-dimensional property space. The number of clusters was found by trial and error, stopping before the number of communities in the smallest cluster with any impact present in more than 80% of its cases became very small. In the Yemen data, this number dropped from 21 for 6 means (i.e. 21 communities assigned to type E) to less than 10 for 7 means.

The above table is the result of considerable reordering of the so-called SPSS final cluster center table:

Final Cluster Centers						
Variable	*Cluster*					
Cluster number given by SPSS	*1*	*2*	*3*	*4*	*5*	*6*
Irrigated farms	0.00	0.00	0.43	0.00	0.07	1.00
Rain-fed farms	0.48	0.27	0.17	0.20	0.14	0.19
Pasture	0.93	1.00	0.91	0.85	0.00	0.95
Wood foraging	0.19	0.20	0.17	0.10	0.02	0.29
Drinking water	0.81	0.00	0.91	0.00	0.02	0.00
Water, other purposes	0.96	0.00	1.00	0.03	0.00	0.10
Housing	0.07	0.01	0.09	0.07	0.16	0.05
Roads to administrative centers	0.15	0.02	0.04	0.08	0.12	0.05
Local roads and trails	0.00	0.00	0.87	1.00	0.00	0.29
The report calls this cluster	C	A	D	B	F	E
# communities in the cluster	*27*	*376*	*23*	*102*	*43*	*21*

[The last two rows were added from outside the SPSS table, the last from cluster member summary in SPSS, then the alphabetic order first by determining the residual cluster (no cells > .80), called "F", then going by decreasing number of communities.]

For tables of this or similar size, the reordering is achieved efficiently by creating a copy, in Excel, with the columns copied one by one in the right order, then by sorting by blockage frequency. Shading is best done manually rather than by conditional formatting.

The final cluster center table, together with the analysis-of-variance (ANOVA) table, should be placed in the methodological appendix of the report. The F-values in the ANOVA table for Yemen clearly show that affected communities were differentiated most strongly by the presence or absence of blocked water sources, a fact borne out both by remarks from community leaders and by correlations with the estimated probability of mine accidents:

ANOVA						
Variable	Cluster Mean Square	df	Error Mean Square	df	F	Sig.
Irrigated farms	4.72	5	0.01	586	327.67	.000
Rain-fed farms	0.52	5	0.18	586	2.82	.016
Pasture	7.77	5	0.03	586	261.45	.000
Wood foraging	0.44	5	0.14	586	3.09	.009
Drinking water	6.74	5	0.01	586	501.67	.000
Water, other purposes	8.68	5	0.01	586	894.63	.000
Housing	0.22	5	0.04	586	5.78	.000
Roads to administrative centers	0.17	5	0.04	586	3.91	.002
Local roads and trails	18.69	5	0.01	586	1588.24	.000

SPSS lets users save the cluster membership of each community. For later convenience, the clusters should be labeled with the same labels as will be used in the report.

However, for future use of cluster analysis, the hierarchical procedure (as opposed to k-means) should be investigated as well. The SPSS manual recommends this for binary data. The hierarchical procedure is conceptually more demanding. The stability of results should be checked through bootstrapping (repeated runs on random samples).

Factor analysis, community background variables

The Impact Survey collects data on several variables that characterize communities, notably population and settlement pattern. Another set of observations aims at establishing the presence of basic institutions and amenities. These variables will vary from country to country and will be decided in the design and pretest phase of each country survey. In Yemen, eight such easy-to-observe institutional variables were included (see below). The number of possible combinations among them is so large that a means to reduce them is needed.

Factor analysis is one of the possibilities. The principal-component analysis of the Yemen data produced a two-factor solution that was relatively straightforward to interpret. The rotated factor loadings suggested that community differences were most significant along the degree of institutional vs. technical modernization.

The factor loadings as such are not of great concern. However, the saved factor scores are important as potential explanatory factor for the behavior of affected communities. The following table presents the key results of the factor analysis. The frequencies of the institutional traits is also shown. The correlations between factor scores and population size are significantly positive, as one would expect by common sense and by almost any theory of community complexity:

VARIABLES AFFECTING INSTITUTIONAL COMPLEXITY

Variable	Percentage Reporting	Rotated component matrix	
		Factor 1	Factor 2
Has secondary school	15%	0.75	0.09
Is ordinary village	90%	-0.72	-0.10
Has health care facility	17%	0.69	0.21
Has primary school	56%	0.59	0.06
Fuel is available	11%	0.47	0.47
Has telephone service	9%	0.21	0.66
Has piped water supply	19%	0.06	0.76
Has electricity	24%	0.06	0.80
Variance explained		36%	16%
Correlation with log of current population		0.48	0.32

Two technical remarks are due:

The information in the database is not always exactly phrased like the variables in the factor analysis. For example, Enumerators report on the number of primary schools in a community, not only on the simple presence of some such schools. Some of the variables, therefore, have to be dichotomized[8]; this is best done in the Excel table before exporting to SPSS.

In some countries, information may be reasonably complete only on a very small number of institutional variables. In such situations, a factor analysis is not possible. In its lieu, a simple additive scale, or a Guttman scale, may be attempted.

Mine accidents and community background

Within the limits of the Impact Survey, the number of recent victims is the only behavioral (as opposed to structural) variable on which communities can be measured with acceptable completeness, reliability and accuracy. Therefore, for good or for bad, it plays a key role in any model that would explain differential success in adapting to the landmine and UXO contamination.

The Yemen analysis related mine accidents to three causal domains – pressure on resources, intensity of past conflict, and institutional endowment. Each concept was broken down into two or several variables for which the survey offered data. Binary logistic regression was used to estimate the influence of these variables on whether the community had been accident-free over the 24 months prior to the survey.

[8] Dichotomous variables, which we use interchangeably with binary variables, have two values only, such as true and false, yes and no. The queries transform the {-1; 0} of checkbox fields, already a binary, into the more straightforward {+1; 0}.

After the Yemen report was done, the model was run again, in a different procedure, zero-inflated Poisson regression, which looks not only at the binary “had some/had no accidents”, but at the number of victims. This procedure is not available in SPSS. We used the Stata software for it.

These were the explanatory variables used. Logarithmic transforms were used when the variable was distributed lognormally:

Concepts	Associated variables	Metric
Pressure on resources	*Size of population*	*Logarithmic*
	Access to water bodies blocked	*Binary*
Intensity of past conflict	*Contaminated area*	*Square meters, logarithmic*
	Distance of nearest mined area to center of community[9]	*Kilometers, logarithmic*
	Years since mines last laid[10]	*Absolute number*
	Distance to nearest (other) community with some recent mine victims	*Kilometers, logarithmic*
Institutional endowment	*Degree of institutional modernization*	*Factor score*
	Degree of technical modernization	*Factor score*

For any model, descriptives of the explanatory variables should be documented in the report appendix. In the Yemen report, this was done for as many as could be fitted on a pagewidth (not very satisfactory!):

[9] Was not used in the binary logistic model. Was used in the zero-inflated Poisson model.
[10] Was not used in the zero-inflated Poisson model. Was used in the binary logistic model.

DESCRIPTIVES OF SOME OF THE EXPLANATORY VARIABLES

	Current population (log 10)	Access to some water source blocked	Estimated mined area surface (log 10)	Distance to nearest mined area (log 10)	Years since mines planted	Distance to nearest town w/ recent victims (log 10)
Number Valid	591	592	592	592	591	592
Number Missing	1	0	0	0	1	0
Mean	2.7491	0.0962	4.9244	2.8515	17.96	0.9396
Median	2.699	0	4.942	2.8588	18	0.8473
Std. Deviation	0.592	0.2952	1.1662	0.4836	9.1	0.5609
Skewness	-0.009	2.744	-0.192	-0.237	0.324	0.389
Kurtosis	-0.04	5.549	0.033	0.858	-0.612	0.126
Minimum	0.78	0	0.85	1.18	1	-0.93
Maximum	4.6	1	7.91	4.39	38	2.78

Logistic regression

The logistic model, as mentioned, uses only part of the information – whether a community had any accidents at all in the period of time considered, or not – it does not use the exact number of victims.

The output, when variable names are written out in full, looks like this:

Variable	*B*	*S.E.*	*Wald*	*df*	*Sig*	*R*	*Exp(B)*
Log10 Distance to nearest other town with recent victims	-0.7508	0.2424	9.5940	1	.0020	-0.1284	0.4720
Technical modernization	-0.4238	0.1554	7.4354	1	.0064	-0.1086	0.6545
Access to some water blocked	0.7376	0.3478	4.4986	1	.0339	0.0736	2.0909
Years since mines were planted last	-0.0305	0.0171	3.1858	1	.0743	-0.0507	0.9700
Log10 Total estimated mined area surface	0.2111	0.1208	3.0543	1	.0805	0.0478	1.2350
Log10 Current Population	0.3990	0.2776	2.0665	1	.1506	0.0120	1.4904
Institutional modernization	0.0786	0.1564	0.2525	1	.6153	0.0000	1.0818
Constant	-3.0405	1.0400	8.5477	1	.0035		
-2 Log Likelihood 425.006							
Goodness of Fit 617.757							
Cox & Snell - R^2 .059							
Nagelkerke - R^2 .109							

Technical notes:

The “B”s are unstandardized. SPSS does not offer standardized coefficients. The odds ratios [Exp(B)], therefore, are misleading. To graph the relative influence of the factors, we used the correlation coefficients [R].

The analysis was done in SPSS version 9. We have not been able to find the option for correlation coefficients in the latest version, 10.

SPSS offers the option to save the probability, for each community with complete data, to have some mine accident. This variable was used, both for selecting high and low-probability communities for community profiles, and for cross-presentation with impact types.

Zero-inflated Poisson regression

This model, sometimes also called “zero-altered Poisson regression”, is particularly appropriate when the zero outcome of the underlying process is qualitatively different from the positive ones. This seems eminently the case with community structures that prevent or favor mine accidents. Essentially, communities can be thought of as belonging to two different regimes. In one regime, the community has the mine danger under control. By definition, the number of accidents is zero. In the other, control is incomplete, and the number of victims follows a Poisson process, which has zero or positive outcomes[11]. The inflated model estimates the influence of the different factors to exclude accidents and to produce a variable number of victims simultaneously.

The results of the Yemen model, run in Stata, were as follows:

***Model component* / variables**	**Coef.**	**Std. Err.**	**z**	**P>|z|**
Number of victims				
Population (Log10)	0.783	0.226	3.462	.001
Distance to nearest mined area (Log10)	-0.526	0.245	-2.146	.032
Estimated surface (Log 10)	0.341	0.119	2.876	.004
Distance to nearest other town with some recent victims (Log10)	0.091	0.211	0.434	.665
Institutional modernization	0.077	0.135	0.569	.569
Technical modernization	-0.026	0.123	-0.213	.831
Some water source blocked	0.369	0.229	1.613	.107
Constant	-2.374	1.061	-2.238	.025
Zero inflation				
Population (Log10)	0.105	0.352	0.299	.765
Distance to nearest mined area (Log10)	-0.417	0.351	-1.186	.236
Estimated surface (Log 10)	-0.064	0.163	-0.392	.695
Distance to nearest other town with some recent victims (Log10)	0.849	0.304	2.793	.005
Institutional modernization	0.003	0.180	0.016	.988
Technical modernization	0.389	0.174	2.233	.026
Some water source blocked	-0.698	0.403	-1.733	.083

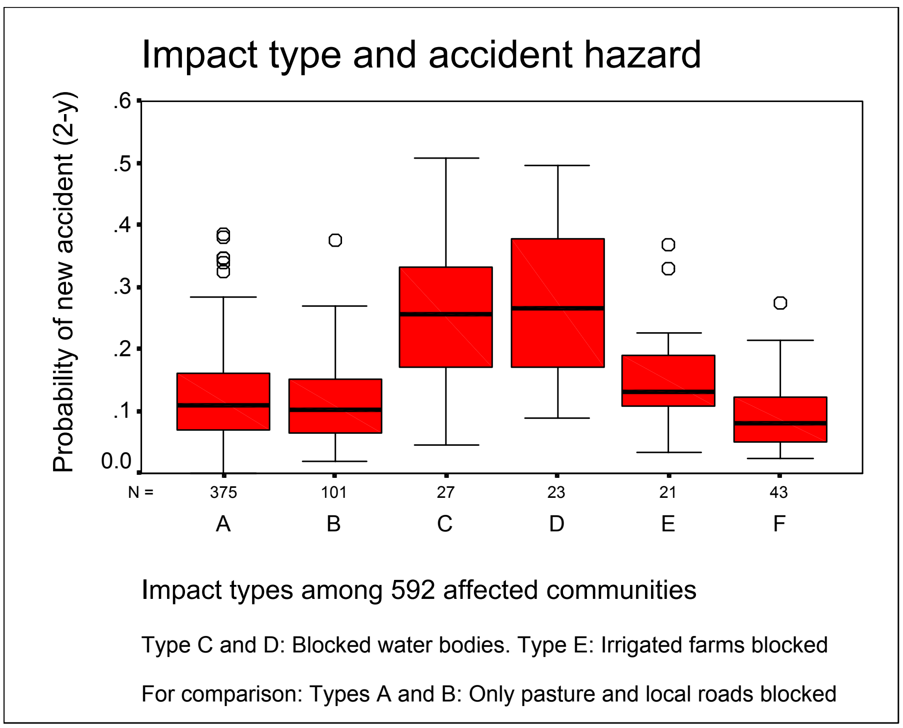
[11] We follow, almost literally, W.H. Greene, Econometric Analysis, Upper Saddle River, Prentice Hall, 2000, pp. 889-893

Constant	1.982	1.462	1.355	.175

The interpretation is obvious: Except for water source blockages, which work on both regimes, the factors exercising significant influence are different from one regime to the other. Several factors that were not significant in the logistic model (population, surface) or were not included (distance to nearest mined area) are significant for the number of victims. Lack of technical modernization and spatial clustering of communities with victims were significant contributors to having some accident in the logistic model and are confirmed as such in the inflation model.

Cluster types vs. accident probability: Box plot, ANOVA

For a better understanding of community adaptation, seeing the correlations between impact combinations and accident probability may be helpful. Box plots can do this very graphically. This is the plot for Yemen:



If the significance of the differences is in doubt, a analysis of variance may be performed.

Statistical support for data editing

The editing function in the Impact Survey takes place in the field, by Field Editors who check for completeness, legibility, missing data and correct skipping of questions. As the Editors gain experience of the mined communities, they will check also for reasonable ranges of variables although this part is nearly impossible to formalize. Another round of editing is done by the data entry personnel, who in theory apply the same criteria, plus the quality of the translation if Editors use coding sheets in a language different from that of the questionnaire.

A number of editing concerns are amenable to statistical methods. Primary among those concerns is the control for

Missing values

Even moderate amounts of missing data in key variables make estimates questionable because the misses are often correlated with particular "real" values or with other variables (e.g., missing estimates for the population of nomadic communities). They also tend to distort, or at least attenuate, the coefficients of multivariate models. Significant missing data in any key variable thus needs to be addressed while data collection is ongoing. For this purpose, in addition to questioning interviewers about items difficult to collect, the pilot test data body needs to be systematically palpitated for blanks.

The function ISBLANK() in Excel is useful in this context[12]. The data table under scrutiny is simply evaluated, cell by cell, on another sheet with identical field names, and the percentage of blank cells is calculated using COUNTIF() and ROWS().

This is an example from one of the country surveys, evaluating the pilot test data pool:

Variable	Missing values
Effects of Local Clearance	93%
Collective Organizations Text	93%
Other Economic Base	93%
Pre-War Population	63%
Total Estimated Surface	*32%*
Current Population	*21%*
Old Victims Killed	20%
Last Conflict Year	11%
Old Victims Injured	9%
Year Mines Last Laid	*8%*
Year Settlement Established	5%
Impact Correction Factor	5%
Recent Victim Assistance	4%
Military Activity	3%
Interview Language	3%
Change of Severity	3%
City Type	3%
Start Date	1%
End Date	1%
Secondary School	1%

[12] A technical hint: isblank() does not recognize blank cells in Excel tables that were copy – pasted from Access. It does work with tables that Access exported in Excel format.

Province GUID	0%
Etc., all others	0%

Many of these are not relevant because they are conditional variables, but the three variables italicized have missing value percentages that are too high and are damaging the survey potential, necessitating correction with the Enumerators and Field Editors. Note that the well-known Pareto principle applies to missing information, with a small part of variables claiming the largest portion of all misses.

Users who prefer to check for missing values in SPSS may do so by going to Analyze – Report – Case summaries, then uncheck "Display cases".

Note that for checkbox variables in IMSMA, there are, by definition, no blanks because it is a two-state box with no null value.

Use of the category "Other"

Frequent use of this residual category is also a warning sign that a validity or interviewer problem is at hand. The frequency can be quickly established using an IF-statement such as

=IF([Data sheet name]![Cell reference]="Other", TRUE, FALSE)

which produced, for the above-mentioned Chad table, the result:

Variable	**Frequency of "Other"**
Economic Base	7%
Medical Facility	3%
Province GUID	0%
Etc., all others	0%

which, in this case, does not signal anything alarming.

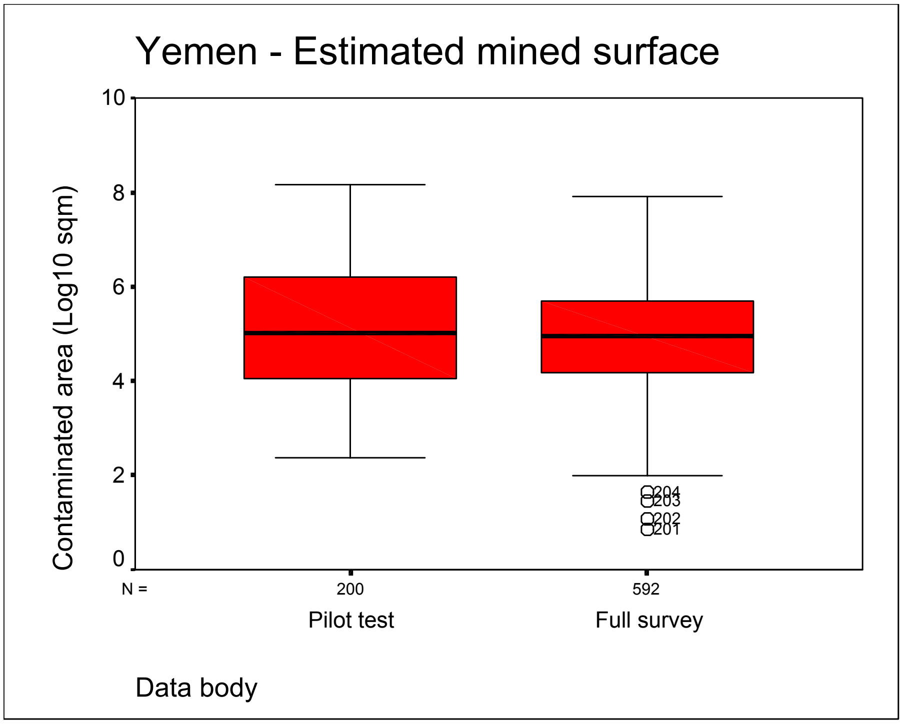
Outliers and inliers

For a limited number of key variables to be selected locally – candidates include population, mined area surface, number of recent victims, years since mines last laid, and others – outlier controls can be applied globally by repeatedly, during data collection already, evaluating extreme values and constructing box plots. Graphic representation of equal-interval or equal-frequency color coded variables in Arc View, using the name of the first enumerator as point label, can also help to detect suspicious patterns. These also apply to inliers, values that fall well within the expected range but are suspiciously uniform[13]. For example, if an Enumerator team returns a dozen

[13] Point stressed by Lyberg, Lars, Daniel Kasprzyk, Some Aspects of Post-Survey Processing, in: Lyberg, Lars, et al. (eds.), Survey Measurement and Process Quality, New York, John Wiley and Sons, 1997, pp. 353-370

surveys of affected communities each with one mined area only, some special verification may seem appropriate.

In Yemen, for example, some of the claims to the largest mined areas from the pilot test were corrected downward, but overall the magnitude range kept increasing, with some communities with very small contaminated areas being recorded during the main data collection. The median stayed remarkably stable:



Scenario development and local indicators

Survey users will want to vary scoring criteria systematically in order to create useful variation for the planning process. In Yemen, government officials convened by survey organization experimented, in very limited measure, with weighting alternatives. Most scenario development that has taken place since the completion of the Impact Survey was about the technical means and cost of mine clearance. Before anything like that is undertaken, the sensitivity of the community categorization to changes in the weights of impact variables should be investigated during the very Impact Survey. Some related questions at the tail end of the survey are also briefly touched upon.

Sensitivity analysis

Impact scores and categories are computed both in the field, by enumerators and their supervisors, and then in the IMSMA Field Module. This is the subject of a different protocol.

IMSMA allows users to change weights and bounds, and also to define a small number of local impact variables in addition to the universally predefined ones. The program then recalculates the scores for all communities. This operation may take several minutes.

For the investigation of numerous scenarios, it is efficient to recreate the scoring formula in an Excel spreadsheet. In a copy of "SAC City S1 MA Victims Master Table.xls", the weights and bounds should be held in a separate worksheet, with names created for each criteria variable using the Insert – Name – Create facility. In the datasheet a column is created for the scores and categories varied under the scenarios, and scores are computed from indicator values in the sheet multiplied with the named weights, such as in:

> =RC[-48]*Mines+RC[-47]*UXO+RC[-46]*Irrigated+RC[-45]*Rainfed+RC[-42]*FixedPasture+RC[-43]*MigratPasture+RC[-39]*WaterDrink+RC[-38]*WaterOther+RC[-41]*LandOther+RC[-37]*Housing+RC[-36]*Roads+RC[-35]*InfraOther+RC[-55]*VictimsRecent

where RC[-x] references the arguments and the self-explanatory terms are the named weighting parameters.

Alternatively, a macro referencing the weights may be used to recreate the scores. The recalculation of the impact category may be entrusted to a custom function.

Country-specific impact variables

Similarly, if a survey creates country-specific impact variables, they should be used in the scoring formula for the scenarios, observing the rules set for them. This particularly regards the fixed weight budget for institutional and resource area blockages.

District-level impact measures

Impact scores are used as the basis for categorizing communities. For practical purposes, the question has come up whether districts can be categorized for mine impact severity, on the basis of community level information. Set-up costs for demining teams may be such that a focus on all high-impact communities regardless of their location may be impractical, and effort may have to be more geographically concentrated.

If districts were simply characterized by the sum of scores of their affected communities, a ranking dilemma may occur. Large, but mildly affected districts may equal, or outrank, small districts with few affected, but high-impact communities. In order to circumvent the dilemma in the Yemen planning process, districts were ranked by the sum of scores of their medium-impact and high-impact communities. The confounding of these two categories, at the district level, was justified with the sensitivity of the score to the number of recent victims, which, as the planning process and clearance move forward, are subject to intertemporal sampling variance.

Technically, the district score table can be produced from the community master table, with a binary created for “Medium or high impact” that goes into the pivot table wizard page area.

Notes on data conversion

Export / import; variable names

(1)We recommend exporting MS Access query results, rather than using copy & paste. For unknown reasons, when a query is copied in full and pasted into an Excel spreadsheet, Excel does not recognize blank cells as such. Therefore, the function “isblank()”, useful for an overview of missing values, does not work. Excel does recognize blanks when the export procedure was used in Access. Access sometimes blocks exporting queries or tables warning that the object is read-only when in fact this option has been unchecked from the file properties. We found that this warning can appear when (a) export is attempted into the same subdirectory, or (b) the same name, except for the file extension, is used. We suggest exporting queries, tables, etc. under different temporary names and into different subdirectories, then moving and renaming them as appropriate. – (2) SPSS can be told to recognize the top row of an Excel table as variable names, and automatically uses them as labels of these variables. However, Excel field names longer than 8 characters or defying SPSS naming conventions are not used for variable names, and these variables are then simply named v1, v2, .., etc. This can lead to confusion when other tables are created with some variables that are substantively different, yet carry the same surrogatory names v1, v2, .., etc. Careful editing would require that all v*-labeled variables be renamed with unique 8-letter names. - (3) A similar difficulty obtains in ArcView. ArcView can import tables from Access and Excel and aliases fields automatically by the original field names. However, the aliases belong to the ArcView *project*, not the underlying tables. In the tables, field names are truncated to the first eight characters, and only these will appear when a table is opened in a different project.

Documenting variables

In the data tables in the IMSMA data pool (= backend), the definitions of user-defined fields (UDFs) and user-defined priority (impact) factors (UDPFs) in the description column of the table design view should be mandatory. Care must be taken also to define the correct data type. For example, some survey data managers have used definable fields to store numeric values, but erroneously defined them as text fields. Similarly, in the Field Module (= frontend), when activating UDFs and/or UDPFs, their definitions should also be written in the description columns of the corresponding tables.

Access has a tool called the Documenter that can help to create the metadata that should accompany any data that is likely to be distributed. The documenter can be found under the *Tools, Analyze, Documenter* menu. Several options are available to help you with your documentation task. Following the above practice of filling out the description column will help in this regard.

In Excel, creating, within the workbook in point, a worksheet that lists and, where necessary, annotates variables is highly recommended. An example is shown in the following screen shot, from Yemen:

	1	2	3	4
1	ColumnNo	Variable	LabelsSPSS	Comments
64	63	Province Center	ProvCent	
65	64	District Center	DistCent	
66	65	SubDistrict Center	SDCent	
67	66	OrdinaryVillage	Village	
68	67	NearestCityWithRecentVictims	CityRVtm	
69	68	Distance	DistNCRV	
70	69	NoCommunitiesWithRecentVictimsWithin10kmRadi	CRV10km	
71	70	HadRecentVictims	HadRcntV	
72	71	Log10CurrentPopulation	LogPop	
73	72	Log10DistanceNearestMinedArea	LogNrMA	
74	73	Log10ContamSurface	LogSurf	
75	74	Log10DistanceNearestCommunityRecentVictims	LogDisNV	
76	75	ClusterMembership	qcl_1	SPSS output

A segment from a list of variables created from a community-level table in Excel. The names were copied into this sheet by using "Paste special – values only, transpose". Variables that belong together substantively are marked by the same color. Some of the variables hold original data, others are calculated. For example, var 63 – 66 hold administrative information from a set of binary variables in IMSMA. 67 – 70, however, hold the results of a macro written to calculate for each affected community the distance to the nearest other community that had some recent victims, excluding itself. Var 75 takes an output from an analysis performed in the statistical application SPSS, which imported data from a reduced version of this Excel spreadsheet. In column 3, the 8-letter or shorter SPSS variable names are put in correspondence.

In SPSS version 10, the so-called "Variable View" makes documentation of variables almost automatic. Care has to be taken, however, to note the meaning of computed variables in their label fields, and to use these labels (minus the spaces) if and when moving any of them back to Excel.

Obtaining query and report templates from SAC

The query and report templates, including two modules for user-defined functions, exist as an empty data pool for MS Access 2000. The pool is designated "P16c_QueryReportTemplates_V3_0_020708.mdb". The version for MS Access 97 is no longer being offered. "P16c_QueryReportTemplates_V3_0_020708.mdb" has been tested so far on a Somaliland pilot test data pool. This pool was created using version 2.2 of the IMSMA Field Module.

Instructions on how to import into the Field Module data pool the queries, modules and reports that P16c offers are given on page 23.

The MS Excel workbook “P16b Data analysis (electronic auxiliary distance calculator) 01 01 18.xls”, which contains the macro to compute distances to the nearest communities with recent victims, is also available.

The templates are included in the collection of electronic protocol files and can be ordered by writing to Peter Harvey <peter@sac-na.org>.