

Lawrence Moulton, Alan Ross¹, Aldo Benini²

Protocol for Assuring High Coverage of Impact Surveys of the Global Landmine Survey

(So-called “Sampling for False Negatives”)



Field Supervisors and Editors in Chad during a tabletop exercise on sampling for false negatives. The table surface is the district. Each card stands for a community; red cards mark those suspected to have mines. White cards, for communities deemed mine-free, are being numbered for a systematic sample draw. “Mines” have been hidden under some of them. Although this chapter is a difficult one in the training of Landmine Impact Survey staff, one of the candidates in this group was already familiar with the concept and practice from an earlier agricultural survey. It is important that by the end of the pilot test country surveys have the procedure in place.

Summary

In order to assure the best possible coverage of mine-affected communities, the Landmine Impact Survey collects expert opinion and evaluates it by three basic rules:

¹ The Johns Hopkins University, School of Hygiene and Public Health

² Survey Action Center

1. All suspected communities are visited. They are composed of true positives, which are then surveyed, and of false positives.
2. During visits, key informants are asked about mine-affected communities nearby. Most of these may already figure in the list of suspected communities; if not, they will also be visited.
3. From the communities deemed mine-free, a sample is drawn; these communities, too are visited. If any of them are found infested, supposedly mine-free communities in their vicinity are also visited.

This protocol is about the logic and practicalities of rule 3. The practice is commonly known as "*sampling for false negatives.*"

For adequate coverage and good comparability among country surveys, sampling plans should assure with 90% confidence that not more than 10% of the non-suspected communities in any affected district are in fact infested. Note that the 10% probability applies to poorly known districts, not to a national average.

If that requirement leads to an impractical sample size, consultation with SAC is urged.

The universe of sampling is the set of all communities deemed mine-free in a given affected district. Sampling is not done in districts on which there is complete consensus that they are mine-free.

For illustration, data on the sampling effort in Yemen is presented. The calculation of coverage is also demonstrated for the Yemen case.

The hurried practitioners may concentrate on the practical "How-to" section (pages 18 - 19). Those in a position to potentially have to explain the procedure and its underlying rationales may wish to assimilate the whole document.

Table of Contents

SUMMARY	1
PURPOSE	4
REVISIONS	4
TERMINOLOGY	5
THE LOGIC OF SAMPLING FOR FALSE NEGATIVES	6
LOT QUALITY ASSURANCE SAMPLING	6
TYPES OF SUSPECTS: AREAS AND COMMUNITIES	7
PROBABILITY AND CONFIDENCE	8
VARIATIONS AMONG DISTRICTS.....	8
SAMPLE SIZE.....	9
SYSTEMATIC SAMPLE.....	10
AREAS OF DISCRETION	10
SAMPLING FRAME: APPROPRIATE DISTRICTS	10
WHEN A FALSE NEGATIVE IS DETECTED	12
APPROPRIATE STAGE OF SURVEY OPERATIONS	13
CASE STUDY: THE CONTROL FOR FALSE NEGATIVES IN YEMEN	14
THE TOTAL SURVEY EFFORT	14
ESTIMATION OF PREVALENCE OF MINE-AFFECTED COMMUNITIES	15
WHAT TO DO PRACTICALLY	16
TRAINING AND TESTING.....	16
FIELD WORK	16
DATA MANAGEMENT	18
FURTHER PERSPECTIVES	19
CONCLUSION	20
APPENDICES	22
SAMPLE SIZE TABLE.....	22
RANDOM NUMBER TABLE	22
EXAMPLE OF A FORM FOR VISITING SAMPLED COMMUNITIES	23

Purpose

The Landmine Impact Survey strives to give a representation as accurate as possible of the number and distribution of mine-affected communities in the surveyed countries. The methodology defines three processes that work together to achieve the desired accuracy: Expert opinion is systematically elicited on communities that are likely affected. All of these suspected communities are then visited. A quality assurance procedure adapted from industrial lot quality control is applied to the non-suspected communities, in order to minimize the risk that affected communities go undetected.

The procedure to follow during fieldwork is simple, but the logic behind it is less than straightforward:

- While the suspected communities are fully enumerated, the non-suspected ones form the only object, within this survey, of a sampling approach³. This calls for special expertise, against the intricacies of which country surveys have to be immunized by simple rules for field work and by outsourcing the analysis.
- Many of the survey users do not understand that there are two, not one, accuracy parameters, in other words, that we cannot absolutely certify that no more than x percent of the non-suspected communities are affected. The solution is to base sample size in an affected district on a probability (10%) as well as on a confidence level (90%) that are standard for the country surveys.
- Against all intuition, the size of the sample does not grow proportionately with the number of non-suspected communities in a district. The sample size quickly moves towards an upper limit, which is not exceeded even in large districts. This has given rise to the “rule of 22” when the above-mentioned probability and confidence parameters are applied.

This note explains the logic of the procedure. It defines rules for fieldwork. It designates some areas in which country surveys have to exercise discretion because of unforeseeable conditions or efficiency concerns.

Revisions

This is version 3 of this document.

The first revision introduced two new elements. Firstly, a rule was given to restrict the area to which the sampling procedure was to apply. If survey staff visited the headquarters of all districts in the country, then those districts producing a complete

³ We are not considering quick overview sample surveys of affected communities, which were planned in the context of a reporting requirement of the Ottawa treaty, but have not so far been attempted anywhere.

consensus that there was no contamination were exempted from the procedure. Secondly, a difference was introduced between the district in which a sample of communities was drawn and the area to search in the event that a sample community turned out affected. For practical reasons, this area had in most cases to be smaller than the district.

The focus of this current revision is mostly semantic. The logic of the procedure was initially stated using a distinction between areas that were definitely landmine-free, those which were probably landmine-free, and those which were probably landmine-affected. Logically that is still correct. In practice, however, the three categories of areas seem to have created confusion in some surveys. Henceforth we operate with a simple distinction between suspected and non-suspected communities within affected districts.

In addition, some practicalities of where and when to conduct sampling in the course of survey operations and the treatment of sampled communities in the database will be discussed. A “how-to-do-it” section is also added.

Terminology

This procedure is for evaluating the existence of communities that have not been suspected by any outsiders as being affected by the presence of landmines, but in fact are affected.

Such a community, in the context of this survey, is also called a “*false negative*”. Only those among the non-suspected communities which are detected to have UXO / mines are “false negatives”. Therefore, it is not correct to talk of a “sample of false negatives” or, even worse, of the set of all non-suspected communities as “the false negatives”.

Correct terms include: “*control for false negatives*” (which includes areas with full enumeration of all non-suspected communities, if that were necessary, as well as the normal sampling approach) and “*sampling for false negatives*”. It is also correct to say that the non-suspected communities are the “potential false negative communities” although this may confuse the casual listener.

Conversely, a community that was suspected to harbor landmines but which upon visit turns out to be mine-free is a “*false positive*”. A “*true positive*” is a suspected community which the survey confirms to be contaminated. A “*true negative*” is a non-suspected community, visited and found mine-free.

Those are terms that staff training in the country surveys must inculcate down to the enumerator level. They must be appropriately translated into the survey language. If a good understanding of what defines those four classes of communities is lacking among field staff, they will not be able to execute the sampling procedure correctly.

In addition there is a small number of terms that the educated survey user may want to know. The impact survey attempts to find a high portion of the affected communities; at

the same time it seeks to avoid wasting resources on communities that are not affected. Two terms refer to those objectives. “*Specificity*” means the probability of finding a negative case if in fact it is negative. “*Sensitivity*” means the probability of finding a positive case if in fact it is positive.

The sampling for false negatives procedure is *sensitive* if it effectively finds many of the affected communities among the non-suspected ones. What survey managers can figure more easily, however, is the reverse of sensitivity and specificity: predictive value. “*Positive predictive value*” is the probability that if we say a case is affected it is in fact affected. At the end of the survey, these are simply calculated by dividing the suspected communities found affected by the number of communities that experts had suspected. Similarly for negative predictive value.

Readers who wish to know more about these concepts, particularly the trade-offs between specificity and sensitivity are referred to a recent article in “*Scientific American*”, “*Better Decisions through Science*”, by John A. Swets, Robyn M. Dawes and John Monahan, October 2000 issue.

The brief game card

- *False negative*: A community that was not suspected, but was found affected
- *False positive*: A community that was suspected, but was found mine-free
- *True negative*: A community that was not suspected and was found mine-free
- *True positive*: A community that was suspected and was found affected

If the expert opinion is perfect, then the survey will report true positives and true negatives only. In this case, the number of true positives will be equal to the number of communities suspected and visited. The number of true negatives will equal the number of communities sampled and visited.

The Logic of Sampling for False Negatives

Lot Quality Assurance Sampling

The general scheme of a Landmine Impact Survey relies on informed opinion as to the location of affected communities. Commonly, the activities pertaining to eliciting and processing those opinions are known as expert opinion collection. Experts, for the impact survey, are usually outsiders to the affected communities. As such, they may be wrong.

There should, therefore, exist quality control mechanisms that will permit some level of verification of the communities that none of the experts designated as mine-affected. Sampling can be used to support a decision rule for areas for which there has not been adequate support for ruling out the existence of landmine-affected communities. The decision rule might be of the form: If, after a random sample, for example, of 25 communities in a suspect area, at least one of the areas has landmines, then commit further resources to investigating all the communities in the designated area. This is the

type of sampling / decision rule utilized for industrial control of manufacturing processes and for immunization coverage in less industrialized countries.

In the current situation, the consequences of making an error in “accepting” the lot, i.e. in deciding that informed opinion was accurate regarding those

communities, is great, and thus we would want to trigger further investigation based on finding even one such community in a given sample. It is important to understand what is rejected when at least one false negative is detected: In industrial lot quality sampling a batch or lot of products are rejected (and literally may be trashed). In this survey, not communities, but pieces of expertise about them, are rejected.

From a lessons-learned site on Lot Quality Assurance Sampling:

“It is important that program managers understand the LQAS method and its limitations. LQAS will not provide precise measures of coverage but will allow program managers to decide whether a population meets certain predetermined standards. It can be difficult to explain the concepts of LQAS to low-level .. staff, although it may not be necessary for them to use the method as a planning tool. Data can be aggregated for several sampling areas to allow for the calculation of overall coverage rates.”

Source: <http://www.basics.org/Publications/tools/immun.htm>

Types of suspects: Areas and communities

Intuitively, when classifying areas of the country as a function of preliminary landmine intelligence gathering, there are at least three levels of designation of geographic areas: areas that are definitely landmine-free; those that are probably landmine-free; those that are probably landmine-affected. Rules can be made for making these categorizations. For example, if an official military source, an NGO working in a district, and the district’s political head all say there are no mines in the district, it might go in the definitely landmine-free category. If the military source says there are no mines, the district’s political head says there are no mines, but a truck driver says they heard of a person being hurt two years ago, then it might go into the probably mine-free category. If the military source says it has landmines, then it might go in the probably landmine-affected category.

Once the intelligence-gathering survey staff have reached down to district headquarters and are interviewing experts there, categories change. They are no longer dealing with areas like provinces or districts; they are aiming at communities, the basic units of the survey. It may still be reasonable to set aside part of the district as definitely landmine-free. This will make sense if the conflict never reached that part. For example, areas in Thai border districts farther than 10 km from the border with Cambodia are commonly believed to be mine-free because hostilities and mine-laying never reached as far inside as that.

However, in the other parts of the district (i.e., probably mine-free plus probably mine-affected), only two categories remain:

- Communities that someone named as suspects of having a landmine problem. Survey staff will visit all of them.
- Communities that have not been named as suspects⁴. All of them are subject to the sampling-for-false-negative procedure, and staff will visit the sampled ones.

The reduction of multi-valued classifications to two-valued ones is, of course, widespread in social systems, such as in the relationship between morals and law, wherever alternatives need to be decided under a binary code (e.g., lawful vs. unlawful). In the case of sampling for false negatives, the reason simply is that a given community can only be inside or outside the set of communities from which a sample must be drawn – there is no third possibility left.

Probability and confidence

Suppose that there are 250 communities in a district, and 10% are landmine-affected. The informed opinion designates 225 communities as being unaffected. If the informed opinion has no value, then about 10% of the 225 communities will be affected. If a random sample of 10 of these 225 communities is taken, and those 10 are visited and interviewed, there is about a 35% chance ($= (1-0.1)^{10}$) that 0 of the 10 will be found to be affected. Thus, there is a 65% chance that, using the decision rule of “at least one triggers further investigation,” it will be decided that there are indeed mine-affected communities in that segment of the district (note that this is tautological, since in fact at least one was found in the sample) and the rest should be investigated. If 22 communities are sampled, then there is a 90% chance of finding at least one affected community in the sample. Another way to think of this is: “If there are 10% or a higher percentage of affected communities among the ones in question, then we will have a 90% chance of deciding upon further investigation if 22 communities are visited.” It will not, however, enable us to certify an area as “mine-free” if no such communities are found in the sample; we have only certified that, with 90% confidence, no more than 10% of the communities are affected.

Of course, these parameters can be varied, and statistical theory will tell the requisite sample size: For example, to say, with 90% confidence, that no more than 5% are affected, we need to sample 45 communities. To have 80% confidence that no more than 5% are affected requires 32 communities.

Variations among districts

These kinds of statements may seem insufficient, but it is necessary to consider the long-run properties of such decision rules. Suppose that in fact, informed opinion is somewhat useful, and that only 5% of the supposedly unaffected communities are affected. Then in the long run (i.e. after doing this in all possible districts), if we use a sample size of 22 communities, only 1.6% of the supposedly unaffected communities will be both affected

⁴ Some readers wanted to know if this definition includes communities that were explicitly named by some experts as mine-free (and, moreover, were not suspected by any others). Yes, it does.

and never visited. This scenario is unrealistic in that it has assumed that in *all* districts, 5% of the supposedly unaffected communities are affected, and we end up visiting 90% of all communities.

In fact, in many districts this percent affected will be 0%, and in some others may be 20%, etc. For example, suppose that in 70% of the affected districts, there are 0% affected communities among those supposedly landmine-free. Further suppose that 20% of the affected districts have 5% affected communities among the non-suspected ones, and 10% of the affected districts have 10% affected communities in the supposedly mine-free group. Overall, that makes for 2% affected communities among the non-suspected ones.

Then the “rule of 22” will end up requiring only 22.5% of the districts to have further investigation (68% of the 20% and 90% of the 10%). In other words, in $22.5/30=75\%$ of the districts with communities affected but not suspected, the procedure actually triggers a search for them. Conversely, in a quarter of the districts where informed opinion was less than perfect, some affected communities will slip through undetected.

This seems like a lot, but it must be put into perspective. If, as assumed in the example, each district has 250 communities, and expert opinion is as good as described, then for 10 such districts, with a total of 2,500 communities, only about 10 affected communities go undetected in the statistical average. In other words, 99.6% of all communities in these districts are either mine-free or are visited. The table summarizes the results:

Table: Example of 10 affected districts with different quality of expert opinion

Proportion affected among non-suspected communities	Districts	Communities	Suspected and visited	Affected among non-suspected	Sampled and visited	Chance of finding at least one affected community	Expected number affected communities not found
0%	7	1,750	175	0	154	0%	0.0
5%	2	500	50	23	44	68% ^(*)	7.4
10%	1	250	25	23	22	90% ^(**)	2.3
Total	10	2,500	250	46	220		9.7

Note: (*) Calculated as $= 1-(1-0.05)^{22}$; (**) Calculated as $= 1-(1-0.1)^{22}$

The ratio of undetected affected communities to all communities, of course, goes down further when both affected and mine-free districts in the country are considered.

Sample size

The number of communities within districts varies. Some districts may enclose a small number of communities only. In smaller districts, to achieve the precision afforded by the “rule of 22” given above, less than 22 communities will be needed. Obviously, if the district has only 20 communities, fewer need to be sampled to give the same degree of

assurance. The following table gives guidelines for the numbers to sample in such situations.

Table: Sample sizes to achieve at least 90% probability of “rejecting” a group of communities if exactly 10% of the communities are affected by landmines.

Number of communities	10	20	30	40	50	60	70	80	90	100	150	200	250	300	1000
Sample Size	9	14	16	17	18	19	19	20	20	20	21	21	21	22	22

Systematic sample

A sampling frame must be available. By this, we mean the agreed-upon list of communities from which to draw the sample. Preferably, a map can be obtained that displays the communities. Then a systematic sampling procedure⁵ can be performed. They are numbered in a geographic order. Then the sampling interval and a random start number are required.

If a map of the communities is not available, but there is a list of all of them, they can be numbered in the list, and a systematic sample similarly drawn; this has the disadvantage that it does not guarantee good geographic distribution.

Areas of Discretion

Sampling frame: Appropriate districts

The expert opinion collection responds to an informational asymmetry. Knowledge of which communities have mines and knowledge of the impacts that the mines have on the polluted communities is generally held by different sets of persons. In order to make the search for affected communities efficient, the survey organization has to proceed in a hierarchical fashion, starting from the national capital. In most situations, the organization will work its way down the administrative tiers. Conceivably, other hierarchical or quasi-hierarchical systems can be explored (such as city nodes-cum-road networks and the associated networks of truck and bus drivers; or the water point of nomads), but practical proposals for such systems have not so far been completed.

At higher administrative tiers – such as regions and provinces -, experts may have general knowledge of broad areas that are mine-affected, but they may not be able to designate

⁵ “Systematic sampling is the selection of sampling units at a fixed interval from a list, starting from a randomly determined point. Compared with random selection, systematic sampling has three advantages: (1) It is easier to perform; (2) it allows easy verification of the selection; and (3) if the list is in some order, the method provides a degree of stratification in respect to the variable on which the list is based. Because of these advantages, systematic selection is much more often used than random selection. In real life, most lists do contain some degree of ordering.” Macro International Inc., Demographic and Health Surveys, Phase III. Sampling Manual. DHS-III Basic Documentation Number 6, Calverton, Maryland, 1996, p. 15.

individual communities. However, it is reasonable to assume that at some lower tier experts will know affected communities by name. The tier at which it is possible to obtain lists of suspected communities or at least clearly localized small clusters of communities we call a “district” in this context. In the local circumstances, this unit may be called differently, for example a sub-district or a canton.

The burden of finding the appropriate tier for the creation of community lists is on the survey organization, and there can be few general rules other than the duty to interview, at the headquarters of affected units at each subsequent tier, several key informants in order to make that determination. It is equally conceivable that in large countries the appropriate tier may differ from one region to another, or from densely settled to sparsely populated areas. The meaning of “district” will then vary in the context of the control for false negatives.

Sampling for false negatives is used only with the communities of an affected district. It is not used within districts for which an adequate set of key informants have agreed that there is no landmine or UXO problem⁶. The justification of this rule is in the variety of contacts that the visiting teams are expected to query, and in the high probability that if the district did have a mine problem, some of the persons contacted at the district headquarters (and, for some districts, already at higher tiers) would know about it.

Similarly, sampling for false negatives is not used for units higher than communities. Country survey managers may decide different rules for the inspection of districts. In countries with poor communications or high social barriers between the survey organization and local experts, it may be reasonable to visit all districts, or at least all the districts within the former conflict zones. In countries where provincial headquarter informants seem knowledgeable about district conditions and exhibit a high degree of commonality in their suspicions of affected districts, it will not be necessary to visit districts deemed mine-free⁷.

The selection of the appropriate administrative tier for the sampling frame has budgetary and timeline implications. Suppose a district is composed of four sub-districts, each with 20 communities deemed mine-free. If the sample is drawn from the 80 communities at the district, the sample size is 20. If samples are drawn from the four subdistricts, the sample size is 4 times 14 = 56. The temptation is obvious to sample from the district. This has been suggested, for example, in Chad, with relation to the “*sous-préfectures*” each of which contains several *cantons*. However, these savings can undermine the validity of the claim to 90% confidence. What is being tested is the expertise that designated certain communities as likely affected, and others, by mirror image, as mine-

⁶ This rule holds also for supposedly mine-free districts in elsewhere contaminated provinces and for supposedly mine-free districts bordering on contaminated districts of neighboring provinces. “Full local consensus” also means that during visits to communities in neighboring districts none of the communities in the district in point was ever designated as mine-infested.

⁷ If later this layer of experts were found in error, with supposedly mine-free districts turning up affected, it would be more straightforward to reject the expertise wholesale and to visit all districts, rather than to attempt sampling for affected districts. The reason is that according to the sample size formula, in a province with 10 districts, 9 will have to be visited, and in one with 20 districts, 14, anyways.

free. Therefore, legitimately, the universe for sampling is the set of all non-suspected communities in the unit where experts significantly designated individual communities.

If that happens at very small units (say, in Chad, at the canton level) and total sample size therefore becomes unviable, it is preferable to settle for a lower confidence level rather than sampling from a “high altitude” with phony confidence claims. Sampling at the level where community-specific expertise was mustered takes into account the variable expert opinion quality across units and justifies relatively small, practical search areas in the event of detecting a false negative.

When a false negative is detected

Communities sampled for the control of false negatives should all be visited. In each of them, the enumerators should ideally have three distinct encounters with local persons during which they establish whether the community has a landmine or UXO problem. If this is the case, they will survey the community.

In addition to surveying the false-negative community, a search is undertaken in the area around it. The rationale for this is that the discovery of a false negative leads the expertise on this and surrounding communities to be rejected, and fresh information has to be sought. Since it is not practical to visit all the so far non-suspected communities of the district, a reasonable search rule has to be defined.

In Yemen, for example, the search area was set as follows: Take the larger of the following two:

- All communities in the district that are within 5 km around,
- The five nearest communities

However, reasonable rules will depend on local circumstances. In addition to the two situations considered in the Yemen survey, it may occasionally be meaningful to include all communities within the area circumscribed by the four other closest sampled-and-found-negative communities (When a community is near an edge of the district, draw lines on the map from the two outermost [widest] of the four nearest villages to the edge, meeting the edge perpendicularly).

A special rule may be needed for the event that during a visit key informants express suspicion about a community in a neighboring district that otherwise is deemed mine-free. Such constellations may be rare, and their complexities are not entirely foreseeable. Such a community should be visited. If it is negative, the search should stop there. If it is indeed infested, a case may arise to reject the entire expertise on that district and to subject it to the lot quality assurance process.

The rules for search areas need to be documented. In any case, when the search produces another false negative, it triggers a new search around this community.

Appropriate stage of survey operations

Ideally, the sample will be drawn, and the communities will be visited, immediately after all communities suspected to be contaminated have been visited, and the surveys have been completed in the district. The justification for doing so at this point in the sequence of survey activities is manifold:

- Visits under the false negative sampling effort will help to further minimize risks of unsurveyed areas, in other words, of failing to find satellite villages for which key informants in central villages claimed to be speaking, but did not actually have knowledge of the mine situation there.
- The effort at this point will improve the survey teams' knowledge of the district for the map that they are supposed to create before finally leaving the district, and will therefore improve the quality of an intermediate product important for quality monitoring and mid-way learning.
- The arrangement facilitates logistics (no need to return to the district) as well as personnel management (if the false-negative situation remained unknown to the end of the survey, demobilization would be more difficult).

There may, however, be reasons why the sampling at that stage of the survey is not practical or not very efficient. Two conditions – one almost the reverse of the other – come to mind that militate for such adaptations:

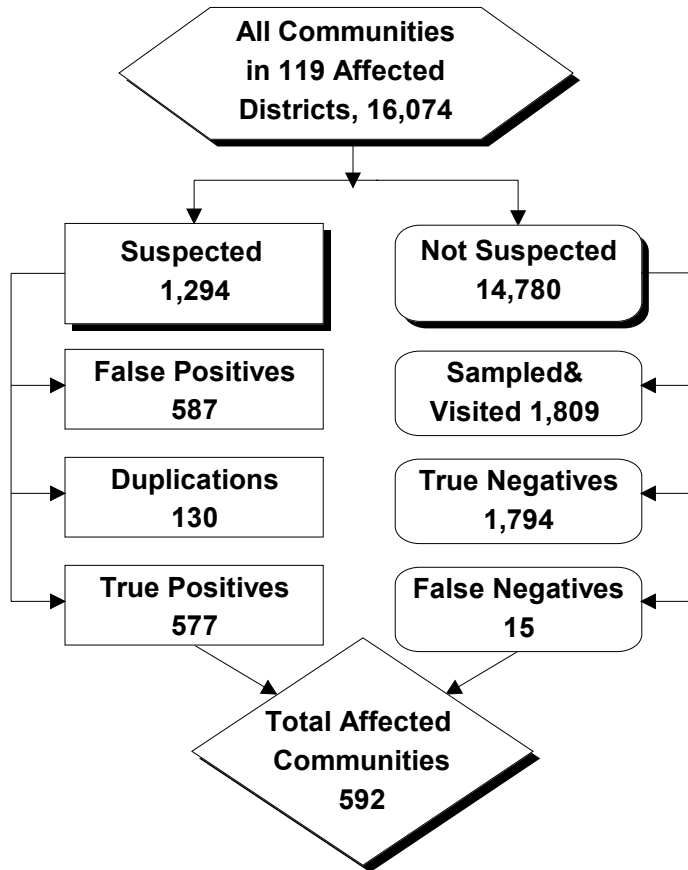
Little is known about the socio-economic environment, and in particular the survey is slow to acquire lists of communities through a government gazetteer or other means. In this situation, it may be difficult to create an adequate sampling frame. In Chad, for example, despite a census report, lists of communities cannot be finalized until field staff meet with the *chefs de canton* who are the traditional authorities knowing the lay of the land.

Contrary to that, in countries where considerable mine action survey work has preceded this survey, lists of supposedly mine-free communities may be construed from existing databases at an early stage already. In Cambodia, for example, specialized false-negative sampling teams finish their work in a given province before the enumerator teams are tasked to survey the suspected communities. However, the sampling teams often run into the problem that existing databases do not record communities with only spot UXO as affected communities. When such a community is visited, it is reckoned to be a false negative, triggering a search around it. This has considerably increased the number of communities to be visited at this stage.

Case Study: The Control for False Negatives in Yemen

The total survey effort

In the Yemen survey, the effort made to visit suspected communities that turned out to be mine-free (false positives) as well as communities sampled for the control of false negatives exceeded the effort required to survey the communities found affected, at least in number, if not in time. The following diagram summarizes the various categories.



The survey effort in Yemen

The figure illustrates how the large number of 16,074 communities in districts known to have a landmine problem was reduced to the much smaller number of 592 communities known, by now, to have such problems. They passed through two separate, but interdependent filters, one using expert opinion, the other using a sampling approach.

These figures underline that expert opinion in Yemen erred chiefly on the conservative side. Experts had a tendency to suspect many communities that in fact were not affected. However, as appears from the sampling and visiting of non-suspected communities, they missed out on comparatively few affected communities. The survey procedures were both highly sensitive (most positive cases were found) and highly specific (most negative cases were found), but the positive predictive value of the expert opinion (the probability that if the expert said the community was affected it was in fact affected) was poor. The negative predictive value cannot really be determined for this opinion because the experts mostly made statements only about suspected communities, not about supposedly mine-free ones.

Based on detailed information of the sampling effort, the prevalence of mine-affected communities was then estimated.

Estimation of prevalence of mine-affected communities

These numbers are survey results from 119 districts.

Of 1294 communities suspected of being affected, 577 were found to be affected. Of the 14,780 “not suspected” communities, 1,809 were sampled (stratified on districts) and investigated, of which 15 were found to be affected, the rest not affected.

Estimation 1:

This estimation procedure assumes that the strategy of going to neighboring communities is equivalent to the full procedure of investigating all communities in a district in which a sampled “not suspected” community was found to be affected. Alternatively stated, it is assumed that because of expected strong spatial correlation, this procedure would have found virtually all affected communities, because of the small chance of having isolated affected communities being randomly (or even haphazardly) distributed throughout a district.

The estimated proportion (prevalence) affected among the 16,074 communities of interest is simply $(577+15)/16,074 = 0.0368$, or almost 4%. Under the above assumption and classical sampling theory, this estimate has zero variance associated with it, since each district in which there were zero affected communities found through the sampling has an unbiased estimate of zero affected and zero variance, and for those with non-zero affected communities in the sample, the assumption is that the search around those found all affected communities, and thus there is no variability.

Estimation 2:

This estimation is carried out under the much milder assumption that the sampled and investigated communities were representative of the non-sampled (and non-suspected) communities in the given district. This results in a “worst-case” scenario, in which we apply the proportion of the sampled communities that are affected to the number of non-sampled communities to estimate the total affected communities in a district.

Let $N = \sum N_h$ be the total number of non-suspected communities from which the samples are drawn, with N_h the number in the h th district. The proportion of affected communities in a sample of n_h in a district is given by: $p_h = a_h / n_h$, and the proportion of affected communities in all the districts, according to the classic formula for stratified sampling (Cochran, *Sampling Techniques*, Third Ed., 1977; Wiley, New York; p.107) is:

$$p_{st} = N^{-1} \sum_{h=1}^{119} N_h p_h. \text{ Further, we estimate its variance by:}$$

$$\hat{V}(p_{st}) = N^{-2} \sum N_h^2 (N_h - n_h) p_h (1 - p_h) / [(N_h - 1)(n_h - 1)].$$

Applying these formulas, we get $p_{st} = 0.001719$, which we apply to $16074 - 1294 = 14780$ to get 25.4 estimated affected communities among the not suspected group (of which 15, or 59%, have been identified). Taking the square root of the estimated variance and using 15 as the lower bound, a 95% confidence interval for the number of affected communities in the not suspected group is (15, 41.4).

The overall estimated prevalence of affected communities among the 16,074 is therefore $(577 + 25.4) / 16,074 = 0.0375$, with a high-end estimate of $(577 + 41.4) / 16,074 = 0.0385$, still less than 4%.

The high degree of sensitivity of the designation of suspected areas resulted in finding very few false negative communities. This means we can be highly confident that almost all affected communities have been found. A somewhat conservative estimate is that only $(25.4 - 15) = 10.4$ affected communities have not been identified as such, which means $(577 + 15) / (577 + 25.4) \times 100\% = 98.3\%$ of all affected communities have been identified.

What to Do Practically

Training and testing

1. Train Field Supervisors, Field Editors, and Enumerators in the protocol.
2. Create auxiliary forms for recording true negatives (and false positives).
3. The second pre-test, if there is one, may, the pilot test must, test the procedure.

Field Work

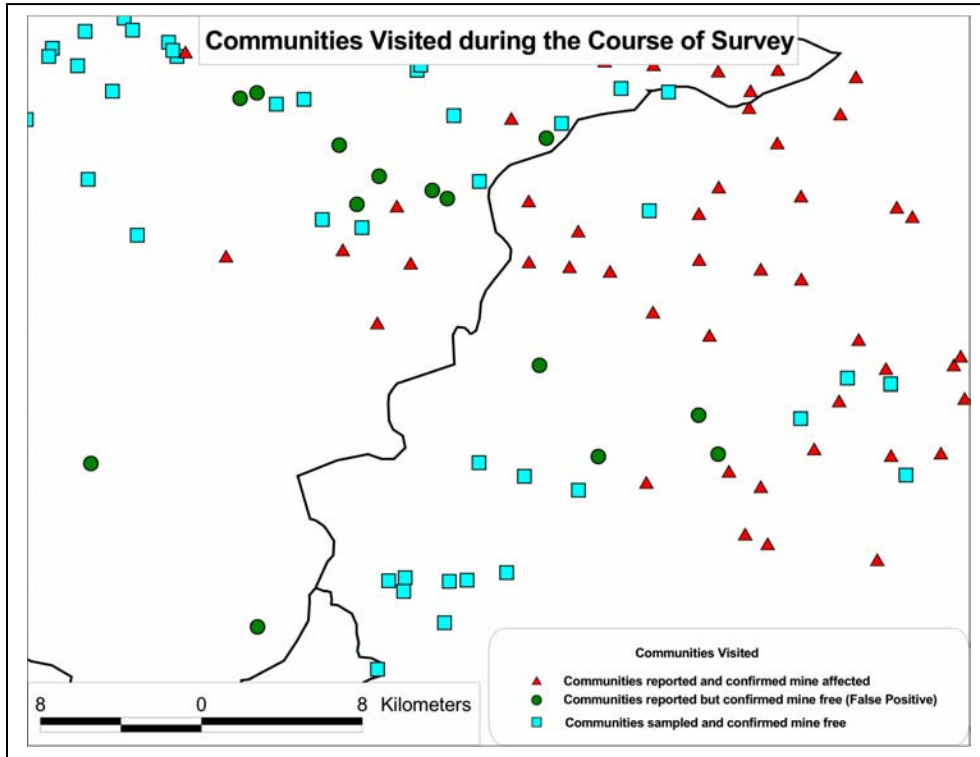
If the sampling in a district can be done by a Field Supervisor, and the sampled communities can be visited by the same enumerator teams who survey the affected communities, proceed like this:

1. Issue field supervisors with
 - district map, gazetteer or census report copies (with locator codes),
 - the list of suspected communities to visit,
 - a random number table or dice,
 - copy of this protocol or at least the sample size table, as well as
 - forms to record sampled and visited communities.

2. Field supervisors are responsible to draw the sample in the district and execute it through their enumerator teams. They draw up a list of all communities in the district⁸ and mark them on the map. For this purpose, gazetteer and map information may have to be updated with the help of local informants.
3. On the basis of that list and of the list of all suspected communities, the complementary list of all communities deemed mine-free is created. This list, as a result of additional information gained from communities already visited, may be smaller than the set that could be constructed using the initial expert opinion.
4. The count of the communities deemed mine-free determines the sample size, which can be looked up in a table. Consecutively number all such communities on the map, or if no map is available, in the list. Divide the count by the sample size and round down to the next integer. This number is the sample interval. Choose the first integer from the random number table equal or smaller than the interval. Number the communities on the map or in the list starting with the random number and increasing by sample intervals.

Example of a sampling interval
If 21 communities are required, and there are 185, then we divide 185 by 21 to get 8.8. Then we round down, and take every 8th community, beginning with a random start within the first 8. This will actually get 23-24 communities to be checked out.
5. Translate the set of selected communities (separately from, or together with, suspected communities to be visited, depending on logistics and work plans) into visiting circuits for enumerator teams. Create a rule for replacing sampled communities that cannot be visited.
6. Enumerator teams will visit these communities. If a local informant confirms that the community is mine-free, seek confirmation in two more, separate, encounters, as when dealing with false positives. Take a GPS reading of the center of the community and collect whatever background information, if any, the country survey format demands on true negatives. Also, ask about mine-affected communities in the area. If any of the sampled communities cannot be visited, record the reasons and how it was replaced if at all. A sample form can be found in the appendix.
7. If the local informants report mines or UXO, survey the community in the same manner as other positives. In addition, do a search for other communities in the area defined by the rules that the country survey applies (see page 12). If the search produces more false negatives, repeat it around each of them.

⁸ Occasionally, this list may not be well definable. Problems may occur when the gazetteer is flooded with poorly defined place names, when there is no gazetteer and local authorities are unable to produce a list, when the list is significantly incomplete, when gazetteer and maps are inconsistent, when a poorly known hierarchical relationship exists between large and small communities, and in other unforeseeable situations. It is not possible to prescribe appropriate adaptations of the protocols for all such situations.



Distribution of affected communities, false positives and true negatives in two districts of the central highland in Yemen. Each of these groups cluster in space. Affected communities (triangles) cluster because of the history of mine laying. False positives (circles) cluster because experts were systematically wrong on certain areas. True negatives (squares) cluster either because the sample was un-systematic, or because the others left only small niches for them to occur at all.

8. At the weekly coordination meeting led by the Field Supervisor, update the district map with the results of the visits of sampled communities.
9. Before leaving the district, complete the map so as to show the status of each community as a result of expert opinion, sampling, visiting and searching. Review the map for large blank areas that may need further investigation and visit some communities there if feasible and necessary.
10. Together with the map, document the effective sample as a list with community names, visiting dates and attached documentation.

If Field Supervisors cannot draw correct samples, or if the sampled communities cannot be visited by the enumerator teams surveying the affected communities in the district, a different procedure will have to be elaborated.

Data management

1. In the administrative section of the community interview questionnaire, the auxiliary forms used to record false positives and true negatives, as well as in IMSMA (using two user-definable fields), create two variables that can be cross-tabulated for the positive/negative status. This is an example:

Prior to visiting, the community was assumed to be:

- *Affected*
- *Possibly affected*
- *Not affected*

When visited, the community turned out to be:

- *Affected*
- *Not affected*

[Later versions of IMSMA may standardize these categories, perhaps in slightly different sets.]

2. When returns on sampled communities arrive, create a record for each true negative at the town level, with other information (e.g. enumerators' names, date of visit) at the survey level and / or at the town level (coordinates belong to both levels) [Similarly, records will be created for false positives.].
3. At certain points of the main data collection phase, for example, 2 – 3 months after the pilot, calculate uptake curves for affected and surveyed, false-positive and true-negative communities and extrapolate the remaining effort to complete data collection against the expert opinion status and reasonable sample size assumptions.
4. During analysis, cross-tabulate communities for negative / positive status and map the four categories. Prepare a district-wise summary, including gazetteer or map counts of all communities to help estimate the prevalence and coverage of mine-affected communities.

Further perspectives

Some questions await further resolution. They primarily concern the storage and the future use of the information on the sampled communities.

- In the Yemen survey, basic information on false positives and on communities sampled for false negatives was kept in spreadsheets, outside the IMSMA base, which is the main survey database. This arrangement was not very productive. At the end of the survey, it took considerable effort to place all the basic statistics on the various categories of communities in one synoptic table. Also, the information kept outside IMSMA is unlikely to be used by mine action planners. It seems desirable that IMSMA users can query the base for communities that the survey may have established as mine-free. In order to do this, two fields will have to be activated in the field module, and rules will have to be set as to how to create records on false-positive as well as sample-visited communities, as shown above (page 18). Such a facility will be particularly helpful in country surveys that can

import electronic gazetteers wholesale. It should become standard in future IMSMA field module versions.

- Secondly, the effort to visit a sampled community, with no further consequences, is smaller than the one required to survey an affected community. Given poor logistics, it is still considerable. Traveling to distant communities and finding (several) local key informants to talk to but on the presence of UXO and mines may seem wasteful. The suggestion comes naturally to piggyback these interviews with some additional information, particularly such as would permit to differentiate affected from unaffected communities. However, one has to keep in mind that expanded information quests in this context would not only slow down the field teams, thereby creating an additional cost, but would, in the eyes of local informants and observers, most likely lack legitimacy. “Why talk to us about community background if we do not have a landmine problem?” Such interviews might return very unreliable information. While these ambitions should not be entirely dismissed and may be pursued on country-specific merit and opportunity, there is an aspect to the greater use of this information that is universally valid. The samples for false-negative control are true probability samples. When excluding the false negatives from them, the executed samples provide sets of communities that were mine-free at the time of the survey. Subsets of these can be used as control groups in future mine action evaluations, thus meeting one of the requirements for valid experimental designs. If additional community background information was collected on the true negatives, close matching with positives may be feasible, adding to the power of the evaluation design. All those are added reasons why the information on false negatives should be stored in the survey main database, in order to be available easily several years after the Landmine Impact Survey.

Both of these considerations point to the necessity to document the sampling for false negatives (and, prior to that, the collection of expert opinion) carefully. While all the rules and procedures are open to review and may need to be creatively adapted to unforeseen local circumstances, documentation of the what and how is a requirement for survey certification.

Conclusion

The Landmine Impact Survey comes into contact with communities for different methodological reasons. Being designated by outside experts and subsequently surveyed by field teams is only one of them. Logically, the various groups of interest are defined by all possible combinations of these properties: (suspected / not suspected) x (visited / not visited) x (surveyed / not surveyed) x (sampled / not sampled) x (searched / not searched). However, such a scheme will only confuse the practitioner, quite apart from the fact that many of these combinations are not possible under the rules of the protocol.

Practically, most of the communities which survey staff visit fall into these categories:

- Suspected, found affected and surveyed (true positives)
- Suspected and found mine-free (false positives)
- Sampled and found mine-free (true negatives)
- Sampled and found affected, surveyed (false negatives)
- Visited during search around false negative, found mine-free (true negative)
- Visited during search around false negative, found affected, surveyed (false negative)

Others include the large group of communities which were neither suspected nor sampled, and which therefore will not normally be visited (some may be visited in order to collect expert opinion). Some of them may in fact be mine-affected. They also include communities that were either suspected or sampled, but were not visited or, if affected, were not surveyed. These latter are anomalies and will normally be few; if their number is considerable, it reflects poorly on the quality of the survey.

While the expert opinion on affected communities may not always follow an easily discernible pattern, the communities visited for the control of false negatives display a clear structure: they form a systematic sample from the set of non-suspected communities within an affected district. As such – like it or not -, they call for the analytical statistician. On the one hand, this tends to pose intellectual challenges for the survey staff, on the other it gives the survey the aura of scientific quality control.

Neither is completely true. Using this protocol, country survey protocols can be written with simple rules that make both the intent and the process understandable to field staff, yet do not demand more than basic arithmetic such as for figuring a sample interval.

The protocol does in fact adapt a scientifically proven quality control procedure from other realms. However, the analogy with industrial Lot Quality Assurance Sampling is less than perfect. What is rejected in the Landmine Impact Survey is not products from well-defined batches, but bits of expertise on communities. To which meaningful lot these communities belong, may be difficult to determine, as the discussion of an appropriate district for the sampling frame proved. Moreover, the practicalities of the survey force a distinction between districts whose communities are sampled and areas that are searched in the event of any community found mine-affected. This is different from the classical LQAS method and takes advantage of the observed regularity that expert errors cluster in space.

Nevertheless, the method is vastly effective in reducing the number of mine-affected communities that go undetected. As such, it helps achieve an important survey objective. The sampling for false negatives may not be feasible to a degree that would achieve confidence levels customary in industrial quality control (where error probabilities as low as one to one billion are not unheard of), but it is important that, through careful training, documentation and dialogue, survey implementers, SAC and Survey Working Group make the best possible use of it in order to produce the compelling results that the mine action community expects.

Appendices***Sample size table***

Probability	.10	.10	.05	.05
Confidence	.90	.80	.90	.80
Number of non-suspected communities in the district	Required sample size			
20	14	11	19	16
40	17	13	28	22
60	19	14	32	25
80	20	14	35	26
100	20	15	37	27
140	21	15	39	29
200	21	15	41	30
240	21	15	41	30
300	22	15	42	30
1000	22	16	44	31

Probability = 10%, confidence = 90% is the standard for Landmine Impact Surveys.

Random number table

50 random numbers between 1 and 50.

44	13	1	24	47
36	11	27	30	41
31	48	30	11	8
41	42	47	3	38
26	44	50	48	6
36	6	8	5	46
4	17	50	5	10
7	39	37	39	6
28	41	19	36	25
9	45	34	26	47

Move from top to bottom, then from left to right, and select the first number smaller or equal the sampling interval and use it as the start number in the systematic sample in the first district. Strike it off and re-use the table in the next district, etc.

Example of a form for visiting sampled communities

This form (header not shown here) is used in the Thailand survey:

IDENTIFICATION

Code: ___/___/___/___
 Province: _____ Subdistrict: _____
 District: _____ Community: _____
 Type of community: Self-defence; Volunteer defence/development; Other; Unknown
 Community reference point: _____
 UTM East: _____; North: _____; Zone: ___; GPS; Map sheet no: _____; Series: _____; Edition: ___

CERTIFICATION DATA COLLECTORS

Data collector 1: Name: _____ Sign: _____ Date: ___/___/200__
 Data collector 2: Name: _____ Sign: _____ Date: ___/___/200__

CHECKED BY

Editor: Name: _____ Sign: _____ Date: ___/___/200__

ENCOUNTER 1

Are there any landmines or UXO in the community? Yes No
 Are there any landmines or UXO in areas used by the villagers? Yes No
 Are there any victims or accidents in the community? Yes No
 Were there mines or UXO in this community in the past? Yes No
 If yes, who cleared them? Local people; Army; Others: _____; When: _____
 Name of informants: _____

ENCOUNTER 2

Are there any landmines or UXO in the community? Yes No
 Are there any landmines or UXO in areas used by the villagers? Yes No
 Are there any victims or accidents in the community? Yes No
 Were there mines or UXO in this community in the past? Yes No
 If yes, who cleared them? Local people; Army; Others: _____; When: _____
 Name of informants: _____

ENCOUNTER 3

Are there any landmines or UXO in the community? Yes No
 Are there any landmines or UXO in areas used by the villagers? Yes No
 Are there any victims or accidents in the community? Yes No
 Were there mines or UXO in this community in the past? Yes No
 If yes, who cleared them? Local people; Army; Others: _____; When: _____
 Name of informants: _____

COMMUNITY STATUS

Prior to visiting, the community was assumed: Affected; Possibly affected; Not affected
 The community turned out to be: Affected; Not affected

As in visits to suspected communities that turn out to be mine-free, the Enumerator must make three distinct encounters. If all the persons met concur that the sampled community is mine and UXO-free, it is considered a true negative.