

# Notes on a design of a simple spatial sampling method (S3M) for assessing coverage of health and nutrition programmes in Liberia

*Valid International*

2018-07-25



# Contents

|  |           |
|--|-----------|
| <b>Simple Spatial Sampling Method (S3M)</b>                                    | <b>5</b>  |
| <b>1 Introduction</b>  | <b>7</b>  |
| <b>2 The survey sample</b>   | <b>9</b>  |
| <b>3 The first stage sample</b>  | <b>11</b> |
| 3.1 Step 1: Find a map . . . . .   | 11        |
| 3.2 Step 2: Decide the area to be represented by each sampling point . . . . . | 17        |
| 3.3 Step 3: Draw a grid over the map . . . . .                                 | 20        |
| 3.4 Step 4: Create an even spread of sampling points . . . . .                 | 22        |
| 3.5 Step 5: Select the communities to sample . . . . .                         | 23        |
| 3.6 Step 6: Label each sampling point . . . . .                                | 25        |
| 3.7 Step 7: Label each triangular tile . . . . .                               | 26        |
| 3.8 Step 8: Describe the sample in numbers . . . . .                           | 27        |
| 3.9 Step 9: Plan data collection . . . . .                                     | 29        |
| <b>4 The second stage sample</b>   | <b>31</b> |
| 4.1 Step 10: Within-community sampling . . . . .                               | 31        |
| 4.2 Step 11: Collect and record the data . . . . .                             | 32        |
| 4.3 Step 12: Enter the collected data . . . . .                                | 33        |
| <b>5 Analysis</b>  | <b>35</b> |



# Simple Spatial Sampling Method (S3M)





# Chapter 1

## Introduction

The Simple Spatial Survey Method (S3M) was developed from the CSAS coverage survey method as a response to the widespread adoption of community management of acute malnutrition (CMAM) by ministries of health. Large-scale programs need a large-scale survey method and S3M was developed to meet that need.

S3M was designed to :

- Be simple enough for MoH, NGO, and UNO personnel without specialist statistical training to perform.
- Provide a general survey method. S3M can be used to survey and map :
- Need for and coverage of selective-entry programs such as CMAM and TSFP as well as universal programs such as EPI, GMP, GFD (general ration), and “blanket” SFP over wide areas.
- Levels of indicators such as those for IYCF, WASH, and period prevalence / cumulative prevalence of ARI, fever, and diarrhoea over wide areas.

This document concentrates on using S3M to assess the need for and coverage of a variety of selective-entry feeding programs. The indicators discussed in this manual are:

- Therapeutic feeding (OTP and TSFP) programs :
- Prevalence of SAM and coverage of treatment of SAM in children aged between 6 and 59 months.
- Prevalence of MAM and coverage of treatment of MAM in children aged between 6 and 59 months.
- Prevalence of MAM and treatment of MAM and in pregnant and lactating women (PLWs).
- Food-based prevention of malnutrition (FBPM) programs :

- Prevalence of need for and coverage of food-based prevention of malnutrition in younger children at risk of developing MAM and SAM.
- Prevalence of need for and coverage of food-based prevention of malnutrition in pregnant and lactating women (PLWs) at risk of developing MAM and SAM.
- Coverage of screening for all of the above programs.
- Coverage of Behaviour Change Communication (BCC) programs focussing on maternal and child health and nutrition to all principal carers of children (usually their mothers) and all PLWs.

# Chapter 2

## The survey sample

The survey method described here uses a two-stage sample:

- **First-stage:** We take an even (or near-even) spatial sample of communities from all of the communities in the survey area.
- **Second-stage:** We take a sample of eligible individuals from each of the communities identified in the first stage of sampling.

Two-stage sampling is used in many survey methods. A typical example of a survey method that uses a two- stage sample is the SMART method that is commonly used for nutritional anthropometry surveys.

The main difference between the sample taken in S3M based surveys and in SMART type surveys is that S3M based samples used a spatial sample in the first stage whereas SMART type surveys use a proportional to population size (PPS) sample.

The advantages of using a spatial first stage sample is that such a sample allows us to identify where (and why) coverage is good, and where (and why) coverage is poor. This information is essential to improving program coverage and ensuring equitable access to services.

A spatial sample can be used to produce equivalent results to a traditional proportional to population size (PPS) sample as is used in (e.g.) SMART type surveys using a weighted analysis. This means that a spatial sample can be made to act as a PPS sample. A PPS type sample cannot, however, be made to act as a spatial sample.



# Chapter 3

## The first stage sample

### 3.1 Step 1: Find a map

The first step in a S3M survey is to find a map of the survey area. A map showing the locations of all towns and villages in the survey area is essential. Try to find a map showing the locations of all towns and villages in the survey area. You may need to update the map to take into account migration and displacement.

For the coverage survey of 2 counties in Liberia, it will be practical and useful to have:

- A small scale-map (a wide area map but with poor detail) of the entire survey area for each of the 2 counties. If the counties are contiguous (i.e., share borders with each other), the small scale map can be of the two counties together. This map does not need to show the location of all towns and villages in the survey area but it gives a general idea of where the 2 counties are located and main towns and locations and roads. Figure 3.1 is a small scale map of Liberia showing counties, roads and main towns and locations. Figure 3.2 is a small scale map of two counties showing all the districts within the county, roads and main towns and locations.



Figure 3.1: Small scale map of Liberia showing counties, roads and points of interest

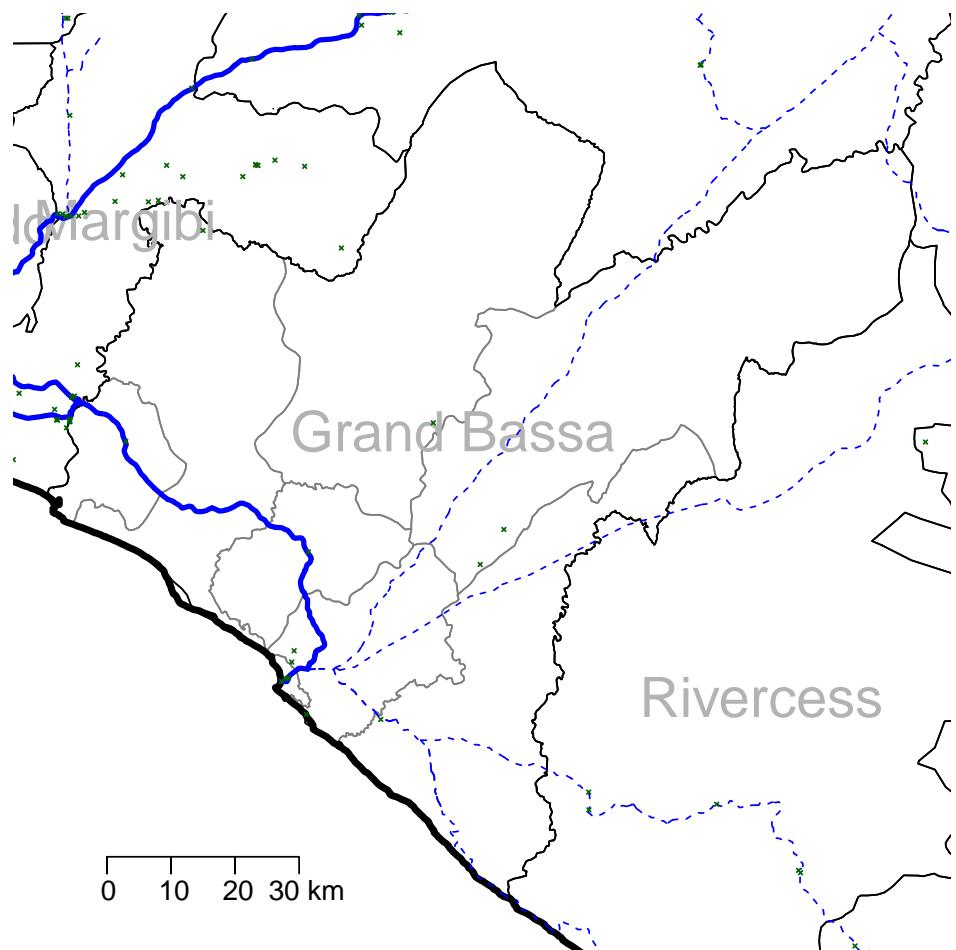


Figure 3.2: Small scale map of Montserrado and Grand Bassa in Liberia showing all districts, roads and points of interest

- A collection of larger scale maps (a small area map but with good detail) of each of the selected counties and each of the districts within those counties in Liberia. Figure 3.3 is a large scale map of Montserrado county showing all districts, roads and all settlements. Figure 3.4 is a collection of large scale maps of each of the districts of Montserrado country showing all roads and all settlements.

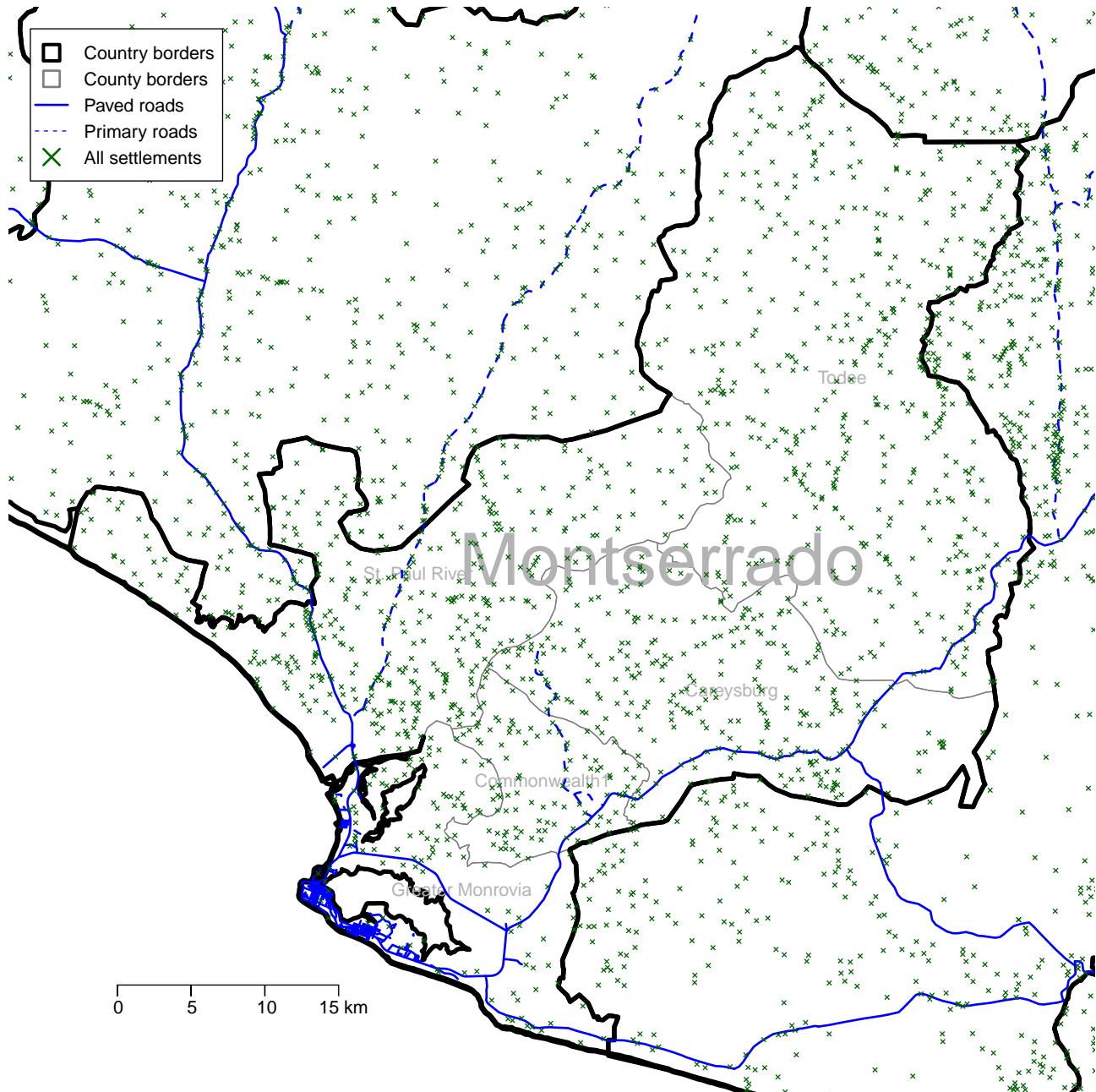


Figure 3.3: Large scale map of Montserrado county in Liberia showing all districts, roads and all settlements (towns, villages)

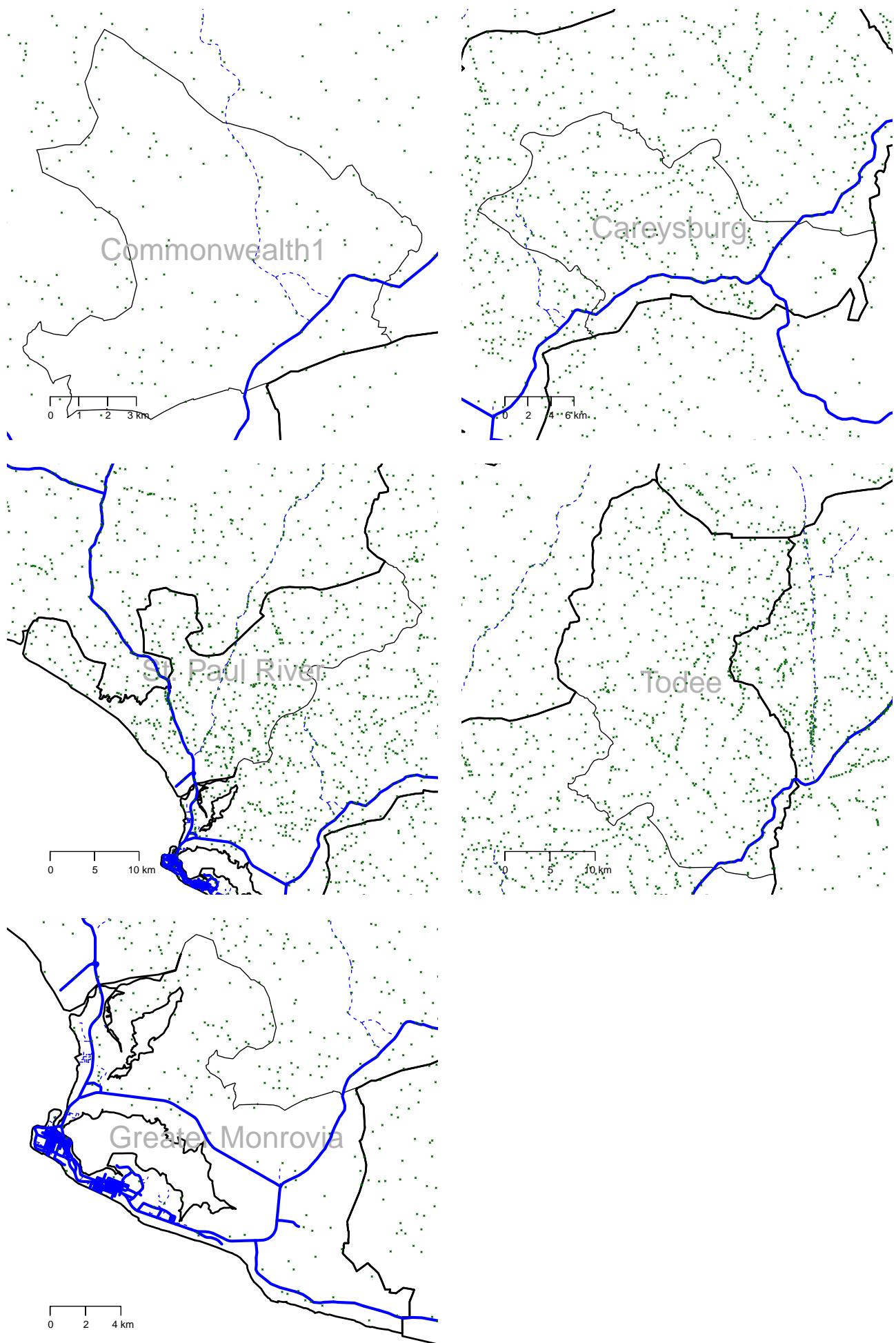


Figure 3.4: Large scale maps of 5 districts of Montserrado county in Liberia showing roads and all settlements (towns, villages)

The small-scale maps in Figures 3.1 and 3.2 will be useful for identifying initial sampling locations.

The large-scale maps in Figures 3.3 and 3.4 will be useful for identifying the precise location of sampling points and for selecting the communities to be sampled.

## 3.2 Step 2: Decide the area to be represented by each sampling point

The easiest way of thinking about this is as a function of the intended maximum distance ( $d$ ) of any community from the nearest sampling point (see Figure 3.5).

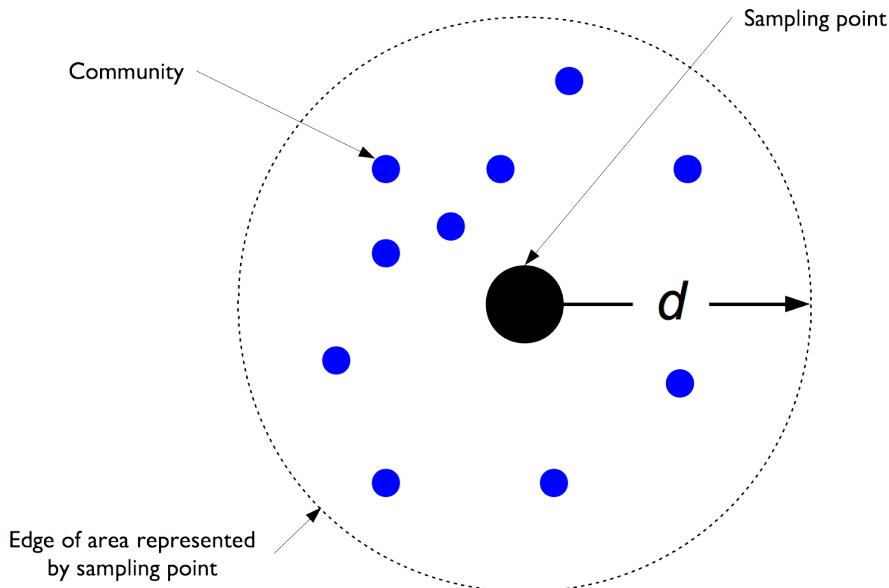


Figure 3.5: Conceptual presentation of the area represented by each sampling point

There are other ways of thinking about  $d$ . These are:

1. **The area of each triangular tile:** This can be calculated using the formula:

$$A = \tan 30^\circ \times \frac{9}{4} d^2$$

For  $d = 10$  km the area of each triangular tile will be about:

$$A = \tan 30^\circ \times \frac{9}{4} d^2 \approx 1.3 \times 100 = 130 \text{ km}^2$$

2. **Practicability:** Most of the time spent in the field when doing a survey will be in travelling to and from sampling points. Having many sampling points can make for an expensive and / or lengthy survey. If you know how many sampling points that you can afford to take ( $m$ ) then you can make a **very approximate** estimate of a suitable value for  $d$  using the following *rule-of-thumb* formula:

$$d \approx \sqrt{\frac{\text{Program Area}}{m}}$$

The value of  $d$  calculated using this formula is approximate and should be used as a starting point for a number of trial samples using the procedure outlined below.

| Pair | Distance | Pair | Distance |
|------|----------|------|----------|
| 1    | 21 km    | 13   | 13 km    |
| 2    | 14 km    | 14   | 11 km    |
| 3    | 13 km    | 15   | 12 km    |
| 4    | 17 km    | 16   | 15 km    |
| 5    | 11 km    | 17   | 13 km    |
| 6    | 14 km    | 18   | 16 km    |
| 7    | 12 km    | 19   | 18 km    |
| 8    | 15 km    | 20   | 13 km    |
| 9    | 16 km    | 21   | 8 km     |
| 10   | 12 km    | 22   | 16 km    |
| 11   | 17 km    | 23   | 18 km    |
| 12   | 14 km    | 24   | 14 km    |

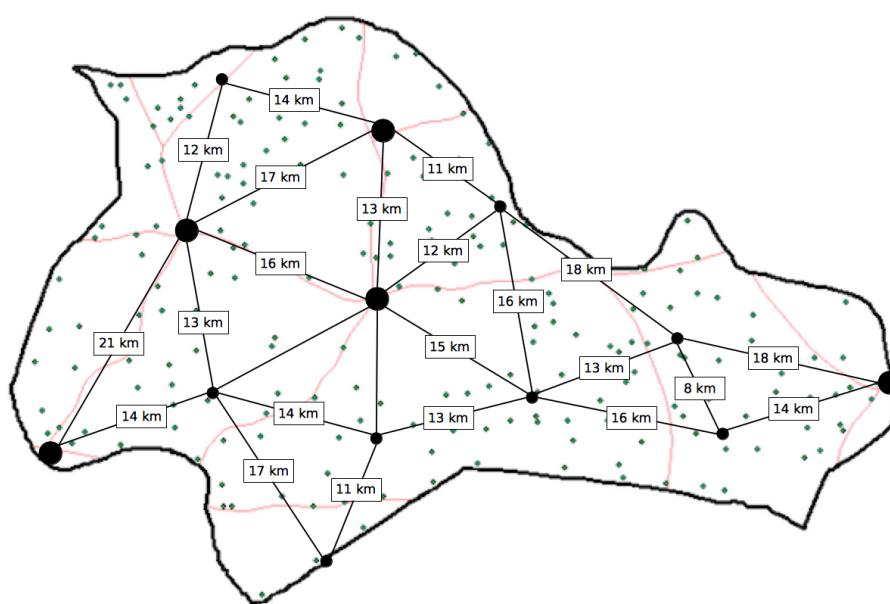


Figure 3.6: Distances of communities to the nearest substantial markets

### Calculations

 Following are steps to estimate value of  $d$  based on distances that carers are willing or able to walk to access services.

Using the information from the distance table, add the distances together:

$$\sum \text{Distance} = 343 \text{ km}$$

Divide the result by the number of paired distances:

$$\frac{\sum \text{Distance}}{\text{Number of paired distances}} = \frac{343}{24} = 14.29 \text{ km}$$

Divide the result by two:

$$d = \frac{14.29}{2} = 7.15 \approx 7 \text{ km}$$

This is an estimate of the distance that carers are willing or able to walk to access services. Only distances between towns and villages with markets are used in this calculation.

A way of deciding a value for  $d$  that is based on the economic geography of the survey area is to set  $d$  to one half of the mean distance between neighbouring pairs of communities with substantial markets.

S3M surveys have been done using a wide range (i.e. from  $d = 8$  km to  $d = 33$  km) of values for  $d$ . A value for  $d$  of 10 km or 12 km will probably be small enough in most circumstances.

### 3.3 Step 3: Draw a grid over the map

The next step is to draw a grid over the map.

The size of the grid is determined by the distance ( $d$ ) that you decided in **Step 2**.

The grid is rectangular rather than square. This allows us to place sampling points at the centres of hexagons in a hexagonal grid without the need to draw a hexagonal grid (see **Step 4**).

The width of the grid in the east-west ( $x$ ) direction is different from the height of the grid in the north-south ( $y$ ) direction.

The width of the grid in the east-west ( $x$ ) direction is calculated using:

$$x = \frac{3d}{2}$$

where  $d$  is the distance ( $d$ ) that you decided in **Step 2**.

The height of the grid in the north-south ( $y$ ) direction can be calculated using:

$$y = \frac{\sqrt{3}d}{2}$$

where  $d$  is the distance ( $d$ ) that you decided in **Step 2**.

For example, if  $d = 10$  km then:

$$x = \frac{3d}{2} = \frac{3 \times 10}{2} = \frac{30}{2} = 15 \text{ km}$$

and:

$$y = \frac{\sqrt{3}d}{2} \approx \frac{1.73 \times 10}{2} \approx \frac{17.3}{2} \approx 8.7 \text{ km}$$

The table below shows the grid sizes for different values of  $d$ :

| <b>d</b> | <b>x</b> | <b>y</b> | <b>d</b> | <b>x</b> | <b>y</b> |
|----------|----------|----------|----------|----------|----------|
| 5        | 7.5      | 4.3      | 13       | 19.5     | 11.3     |
| 6        | 9.0      | 5.2      | 14       | 21.0     | 12.1     |
| 7        | 10.5     | 6.1      | 15       | 22.5     | 13.0     |
| 8        | 12.0     | 6.9      | 16       | 24.0     | 13.9     |
| 9        | 13.5     | 7.8      | 17       | 25.5     | 14.7     |
| 10       | 15.0     | 8.7      | 18       | 27.0     | 15.6     |
| 11       | 16.5     | 19.5     | 19       | 28.5     | 16.5     |
| 12       | 18.0     | 10.4     | 20       | 30.0     | 17.3     |

When drawing the grid make sure that it covers the entire survey area.

It is usually best to draw a grid that covers an area that is a little larger than the entire survey area. This helps to ensure that the survey will sample from the entire survey area.

The grid can be drawn using marker pens onto plastic film overlaying the map. This protects the map and allows you to reposition the grid to improve the coverage of the sample should this be needed.

If you are drawing the grid directly onto the map then use a soft pencil (e.g. a 2B or #1 pencil). A soft pencil will not damage the surface of the map and is easy to erase using a soft rubber eraser should you make a mistake or need to draw a different grid.

### 3.4 Step 4: Create an even spread of sampling points

Sampling points are located at the intersections of the rectangular grid in a staggered fashion. Alternate intersections of the grid in the x (east-west) and y (north-south) directions are used :

Note how this process places sampling points at the centres of hexagons in a hexagonal grid without the need to draw a hexagonal grid. Make sure that your sample points go right to the edge (or even over the edge) of the survey area. This helps to ensure that the survey will sample from the entire survey area.

## 3.5 Step 5: Select the communities to sample

Select the community (or communities) closest to the sampling points identified in **Step 4**.

The position of the sampling point is moved to the position of the selected community. This is shown in the diagram above.

You may drop sampling points if you find that many sampling points are clustered closely together.

You may move or add sampling points if you find that there are populated areas that do not contain sampling points.

The aim is to create a roughly even spread of sampling points over the entire survey area.

### Caution

**⚠** The S3M sample is defined using a systematic sampling method. Like any systematic sampling method, an S3M sample can produce biased estimates if there is periodic variation in prevalence and / or coverage and the sampling points tend to coincide with this periodicity. This is difficult to control for without prior knowledge of the periodic variation, although simple checks such as ensuring that sampling points are not all in valleys or all on hilltops, and adjusting the grid position accordingly, should help to minimise this problem.

A good way to check if you have an even spread of sampling points over the entire survey area is to do a trial *triangulation* of the selected sampling points. This involves dividing up the survey area into non-overlapping triangles with a sampling point at each vertex.

There will usually be many ways to divide the survey area into triangles. The best triangulation is one that results in small equilateral triangles (i.e. triangles with all sides of equal length) or small and nearly equilateral triangles. Avoid long and narrow triangles. Avoid large triangles.

You can triangulate “by eye” or automatically (i.e. using a computer). If you use a computer to do this then you should use software that produces a *Delaunay triangulation*.

You may drop sampling points if you find that many sampling points are clustered closely together.

You may move or add sampling points if you find that there are populated areas that do not contain sampling points.

The aim is to create a roughly even spread of sampling points over the entire survey area.

The diagrams above show a trial triangulation with only a few long and narrow triangles :

- One sampling point (labelled x1) has been added to ensure that the sample covers almost the entire survey area.

- Four sampling points (labelled m1, m2, m3, and m4) have been moved to ensure that there are few long and narrow triangles.

**The sample will, to some extent, be dictated by the distribution of communities in the survey area. It is usual to find that you have some large triangles and some long and narrow triangles in your final triangulation. You should try to keep the number of these “problem” triangles to a minimum.**

The process of selecting communities to sample is:

1. Start by defining the sample using the grid based approach outlined above.
2. Use a trial triangulation. This can be done “by eye” or using a computer. Check for an even spatial sample:
  - Most triangles should be short and wide.
  - Very few triangles should be long and narrow.
  - The triangles should be of roughly equal size.
  - The complete set of triangles should cover all (or almost all) of the survey area.
3. Move or add sampling points to improve the sample (i.e. to avoid long and narrow triangles, to avoid large triangles, to make triangles roughly equal in size, and to ensure the sample covers all or almost all of the survey area). Triangulate again. Repeat this process until you are happy with the sample.

**The sample will, to some extent, be dictated by the distribution of communities in the survey area. It is usual to find that you have some large triangles and some long and narrow triangles in your final triangulation. You should try to keep the number of these “problem” triangles to a minimum.**

## 3.6 Step 6: Label each sampling point

Give each sampling point a unique identifying label:

- The label may be a number or a name.
- The label must be unique.
- The label is used to identify which community belongs to which sampling point.

These labels are used when collecting, organising, and analysing data.

### 3.7 Step 7: Label each triangular tile

Give each triangular tile a unique identifying label:

- The label may be a number or a name.
- The label must be unique.
- The label is used to identify which communities belong to which triangular tile.

These labels are used when collecting, organising, and analysing data.

#### Note

 This step is only required if you will be using the simple analysis provided by the RAnalyticFlow workflow. More complicated analysis methods will usually not use the tile labels created in this step.

## 3.8 Step 8: Describe the sample in numbers

To analyse and map data you will need to create two files:

- **The PSU file:** This is a file that describes each sampling point (a sampling point is sometimes called a primary sampling unit or PSU). This file **must** have the following named columns and the names **must** be **exactly** as given here:

| Column Name       | Contains   | Notes   |
|-------------------|--|---|
| <b>psu</b>        | Unique identifying label for a sampling point (from <b>Step 6</b> ). | Must be unique.   |
| <b>pop</b>        | Population of the community at the sampling point.                   | Can be taken from census or program data or collected when the sampling point is visited. |
| <b>latDegrees</b> | Latitude of the sampling point                                       | Optional  |
| <b>latMinutes</b> | Minutes of latitude should be recorded in decimal minutes            | Optional  |
| <b>lonDegrees</b> | Longitude of the sampling point                                      | Optional  |
| <b>lonMinutes</b> | Minutes of longitude should be recorded in decimal minutes           | Optional  |

The PSU file may also contain other columns such as district name, village name, and date visited. These columns may help when describing the sample in survey reports. If (e.g.) the survey area covers many districts then a column giving the district name will be useful if district-level summaries are required.

- **The TILE file:** This is a file that links sampling points to tiles. This file has only four columns which must have the exact names:

| Column Name | Contains  |
|-------------|---|
| <b>tile</b> | Unique identifying label a tile (from <b>Step 7</b> ) |
| <b>psu1</b> | Identifying label for vertex 1 of the tile.           |
| <b>psu2</b> | Identifying label for vertex 2 of the tile.           |
| <b>psu3</b> | Identifying label for vertex 3 of the tile.           |

For example, the first four rows of the TILE file for the sample shown in the illustration to **Step 7** would be:

| tile | psu1 | psu2 | psu3 |
|------|------|------|------|
| 1    | 4    | 5    | 8    |
| 2    | 5    | 8    | 10   |

| tile | psu1 | psu2 | psu3 |
|------|------|------|------|
| 3    | 5    | 9    | 10   |

The TILE file is only required if you will be using the simple analysis provided by the **RAnalyticFlow** workflow. More complicated analysis methods may not need a TILE file.

Both of these files can be created using a spreadsheet package such as Microsoft Excel or OpenOffice Calc but they should be exported as comma-separated-value (CSV) files before use by the **RAnalyticFlow** workflow. Do **not** use commas (i.e. “,”) in any cell when entering data.

## 3.9 Step 9: Plan data collection

Now you have identified the location and number of sampling points you can draw up a project plan (using your favourite or institutional project management tools) and budget. When drawing up the timetable and budget it is common to assume that one team will sample one sampling point in one day.



# Chapter 4

## The second stage sample

### 4.1 Step 10: Within-community sampling

The sampling process that you use to select a sample from a community will depend on what the survey is investigating.

If you are investigating multiple indicators which apply to different groups of individuals then you may find it easier to use different sampling methods for different indicators. You can think of this as having different surveys for different indicators sampled from the same set of communities at the same time.

If you are evaluating a program treating SAM you can efficiently find SAM cases using an active and adaptive (snowball) case-finding method. If you are also evaluating TSFP and FBPM programs using house- to-house and door-to-door screening then you will not need to use active and adaptive case-finding as the screening process will also identify SAM cases. SAM is a rare condition. This means that estimates and maps of SAM prevalence and SAM treatment coverage may be based on only a small number of cases from a small number of sampling points. This problem can be avoided by using a “top-up” sample collected using an active and adaptive case-finding method in a small set of communities neighbouring the selected sampling point.

If you are evaluating (e.g.) TSFP and FBPM programs then the active and adaptive case-finding may only find the most severe cases. In this case, house-to-house and door-to-door screening will be the most appropriate way to find eligible persons (cases). If a sampling point is located in a village then the entire population can be screened. If a sampling point is located in a large village or town then the village or town can be divided into a set of segments and the entire population of only a sample of these segments are then screened (see **Appendix A**).

If you are evaluating a BCC program then you can apply your BCC questions to the carers of children and to the PLWs that you find in the sample.

## 4.2 Step 11: Collect and record the data

Datasets for need and coverage surveys of selective-entry feeding programs are usually simple and can be collected using simple tabular forms.

Here is an example of a form used to collect coverage data on TSFP and FPBM programs for children:

Here is an example of a form used to collect coverage data on TSFP and FPBM programs for PLWs:

The “BCC confirmed by message recall” column records whether the respondent can recall any messages that match with the program’s BCC key messages.

The data collected using the tabular forms will allow you to estimate or classify need and coverage. They do not, however, allow you to know the reasons for coverage failure. To collect this data we apply a “barriers” questionnaire to uncovered cases (PLWs) or the carers of uncovered cases (children). Here is an example of a barriers questionnaire:

The barriers questionnaire is applied only to non-covered cases.

Remember to refer all non-covered cases to the appropriate program.

## 4.3 Step 12: Enter the collected data

Coverage survey data is simple enough to be entered (and analysed) using spreadsheet packages such as Microsoft Excel or OpenOffice Calc.

If you want to analyse your data using a spreadsheet then you can use column names that make sense to you. If you want to analyse your data using the **RAnalyticFlow** workflow (see **Step 13**) then the column names **must** be exactly:

### Children

| Column            | Contains                                       |
|-------------------|--|
| <b>age</b>        | Age of child in months                         |
| <b>sex</b>        | Sex of child 1 = Male; 2 = Female              |
| <b>muac</b>       | MUAC in <b>mm</b>                              |
| <b>screened</b>   | MUAC taken previously (in previous<br>4 weeks) |
| <b>oedema</b>     | Oedema   |
| <b>program</b>    | Program child is attending                     |
| <b>confirmBCC</b> | BCC contact (confirmed by message<br>recall)   |



# **Chapter 5**

## **Analysis**