

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

Valid International

2018-08-08

Contents

Simple Spatial Sampling Method (S3M)	5
1 Introduction	7
2 The survey sample	9
2.1 The first stage sample	9
2.2 The second stage sample	36
2.3 Sample size considerations	39
2.4 Urban and rural sample considerations	41
2.5 Overall sampling frame	41
2.6 Estimated time to complete full survey	44
3 Indicators	45
3.1 CMAM coverage	45
3.2 Vitamin A supplementation	51
3.3 Iron-folic acid (IFA) supplementation for pregnant women	52
3.4 Micronutrient powder supplementation	52
3.5 IYCF counselling	53
4 Questionnaire	55
4.1 CMAM coverage survey instruments	55
4.2 Survey for children 6-59 months and their mothers	58
5 Data management and analysis	63
5.1 Analytical approach for estimating coverage indicators	64
5.2 Analytical approach for mapping coverage indicators	67

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:

- Treatment of SAM in children aged between 6 and 59 months;
- Vitamin A supplementation;
- Micronutrient powder supplementation;
- Ferrous sulphate-folic acid supplementation; and,
- Infant and young child feeding counselling.

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 use 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.

2.1 The first stage sample

2.1.1 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 2.1 is a small scale map of Liberia showing counties, roads and main towns and locations. Figure 2.2 is a small scale map of two counties showing all the districts within the county, roads and main towns and locations.



Figure 2.1: Small scale map of Liberia showing counties, roads and points of interest

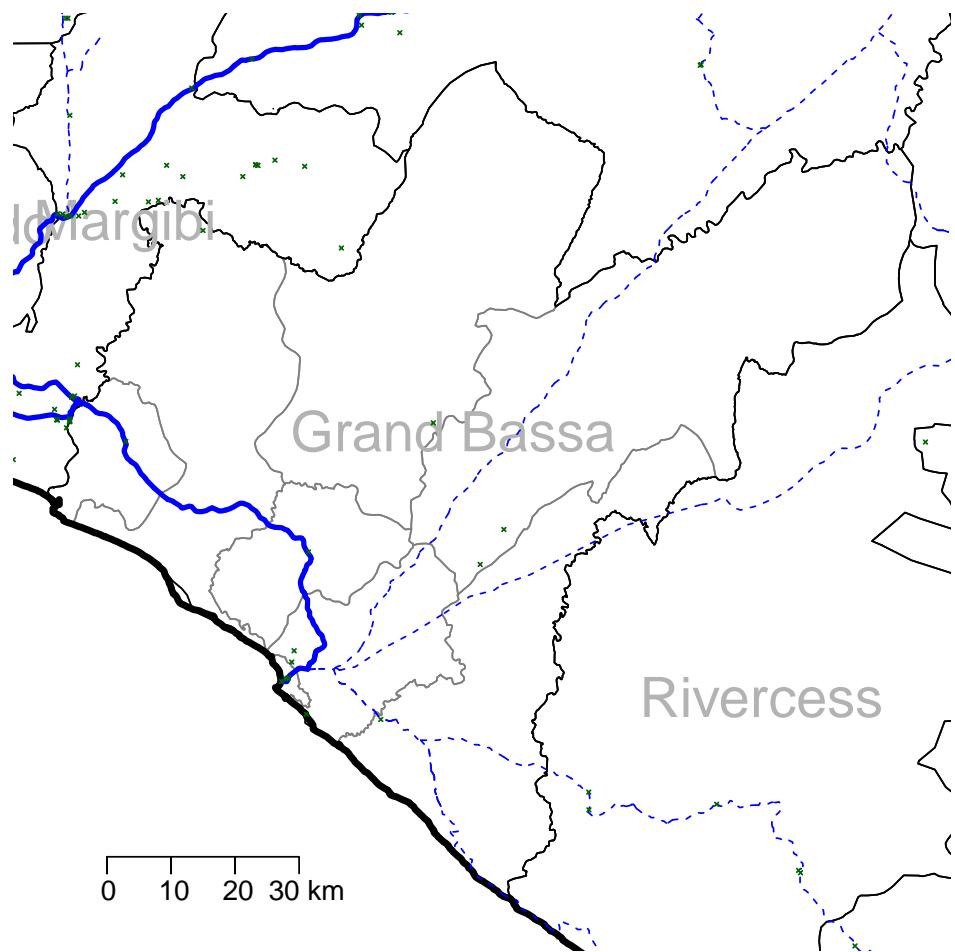


Figure 2.2: Small scale map of Montserrado and Grand Bassa in Liberia showing all districts, roads and points of interest

- 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 2.3 is a large scale map of Montserrado county showing all districts, roads and all settlements. Figure 2.4 is a collection of large scale maps of each of the districts of Montserrado country showing all roads and all settlements.

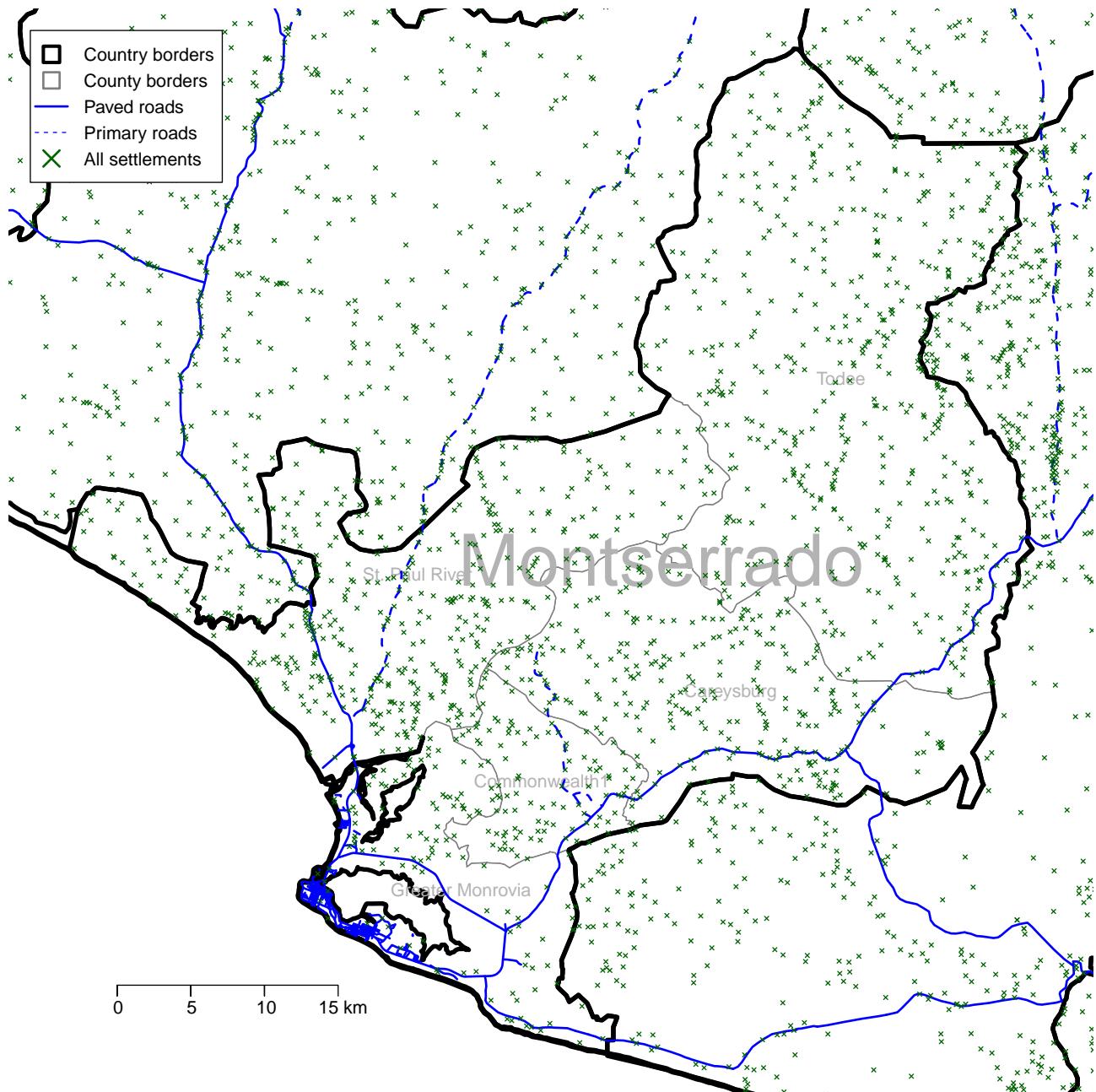


Figure 2.3: Large scale map of Montserrado county in Liberia showing all districts, roads and all settlements (towns, villages)



Figure 2.4: Large scale maps of 5 districts of Montserrado county in Liberia showing roads and all settlements (towns, villages)



Figure 2.5: Large scale map of Grand Bassa county in Liberia showing all districts, roads and all settlements (towns, villages)



Figure 2.6: Large scale maps of 8 districts of Grand Bassa county in Liberia showing roads and all settlements (towns, villages)



Figure 2.7: Large scale maps of 8 districts of Grand Bassa county in Liberia showing roads and all settlements (towns, villages) continued

The small-scale maps in Figures 2.1 and 2.2 will be useful for identifying initial sampling locations.

The large-scale maps in Figures 2.3, 2.5, 2.4, 2.6 and 2.7 will be useful for identifying the precise location of sampling points and for selecting the communities to be sampled.

2.1.2 Step 2: Decide the area to represent 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 2.8).



Figure 2.8: 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

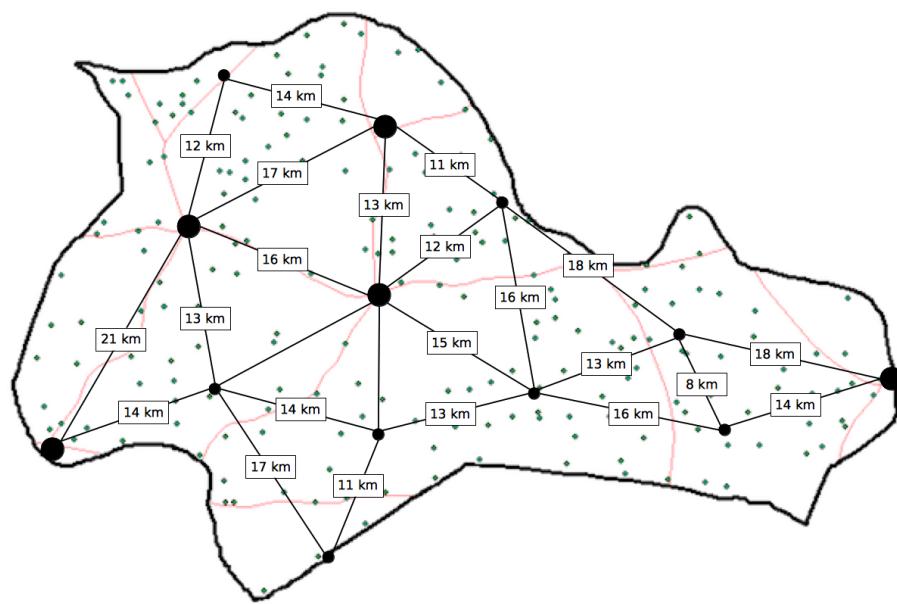


Figure 2.9: Distances of communities to the nearest substantial markets

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.

2.1.3 Step 3: Draw a grid over the map

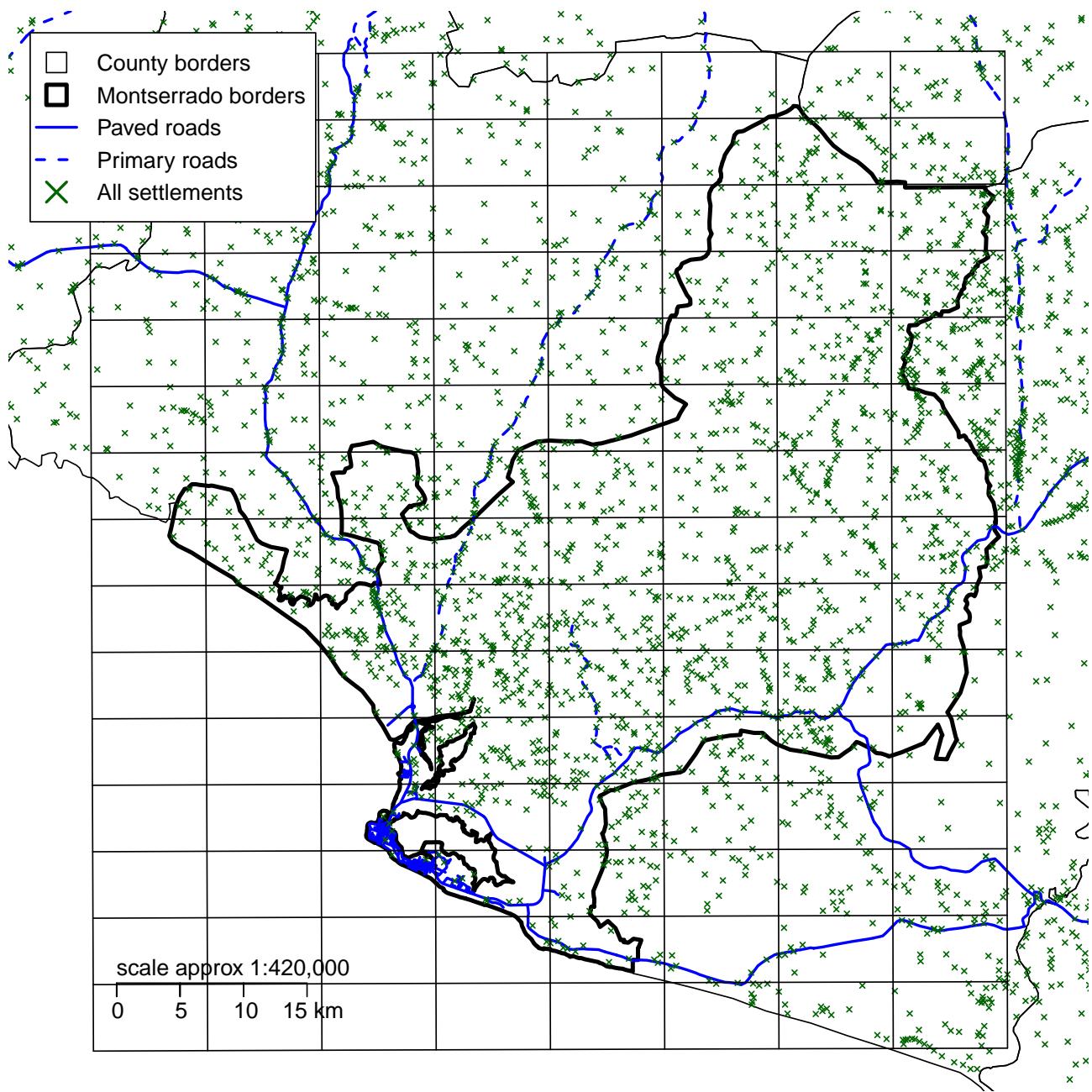


Figure 2.10: Montserrado county with a rectangular grid defined by d of 6 km



Figure 2.11: Grand Bassa county with a rectangular grid defined by d of 10 km

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, in Figure 2.10, we used $d = 6$ km. This value of d creates a rectangular grid with the following dimensions:

$$x = \frac{3d}{2} = \frac{3 \times 6}{2} = \frac{18}{2} = 9 \text{ km}$$

and:

$$y = \frac{\sqrt{3}d}{2} \approx \frac{1.73 \times 6}{2} \approx \frac{10.38}{2} \approx 5.2 \text{ km}$$

So, the grid in Figure 2.10 is 9 km long on the east-west direction and 5.2 km on the north and south direction.

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

d	x	y	d	x	y
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.

2.1.4 Step 4: Create an even spread of sampling points



Figure 2.12: Montserrado county with a rectangular grid defined by d of 6 km and alternating intersections of the grid used to identify sampling points

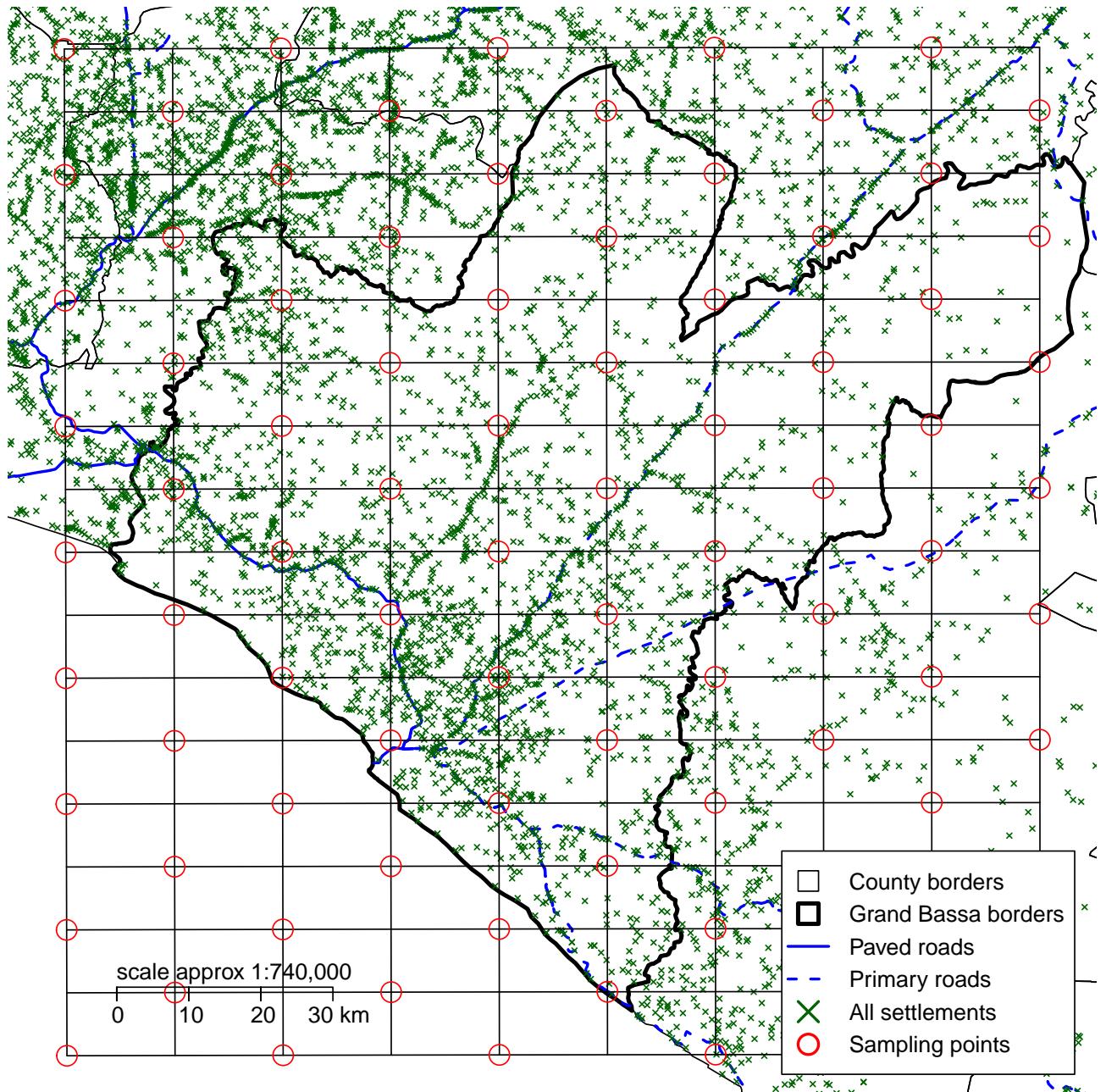


Figure 2.13: Grand Bassa county with a rectangular grid defined by d of 10 km and alternating intersections of the grid used to identify sampling points

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:



Figure 2.14: Selecting alternating intersections of the grid in the x and y directions to spread sampling points evenly

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.

2.1.5 Step 5: Select the communities to sample

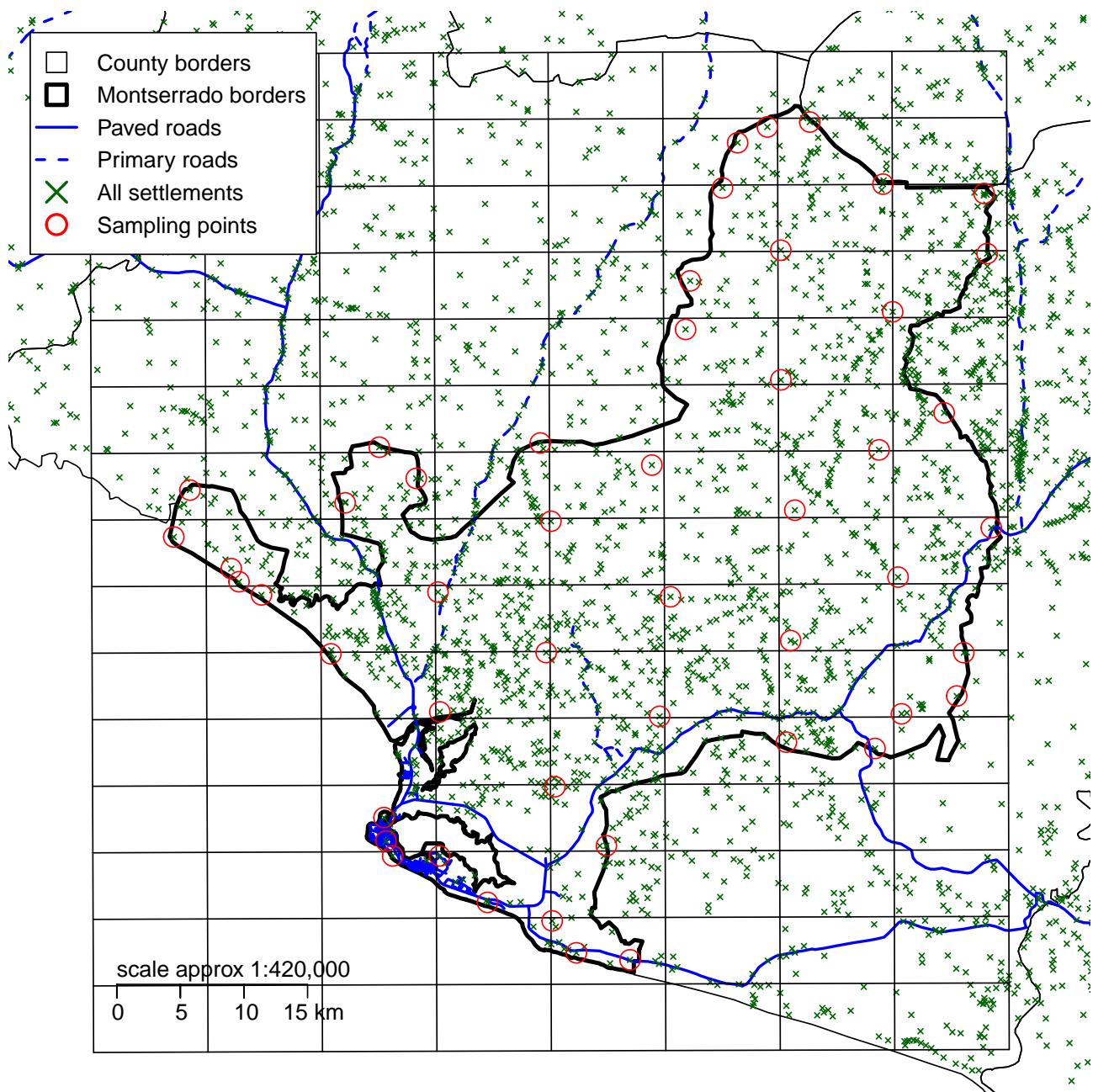


Figure 2.15: Montserrado county with a rectangular grid defined by d of 6 km and sampling points moved to the nearest communities



Figure 2.16: Grand Bassa county with a rectangular grid defined by d of 10 km and sampling points moved to the nearest communities

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.



Figure 2.17: Montserrado county with a rectangular grid defined by d of 6 km and sampling points moved to the nearest communities showing test triangulation

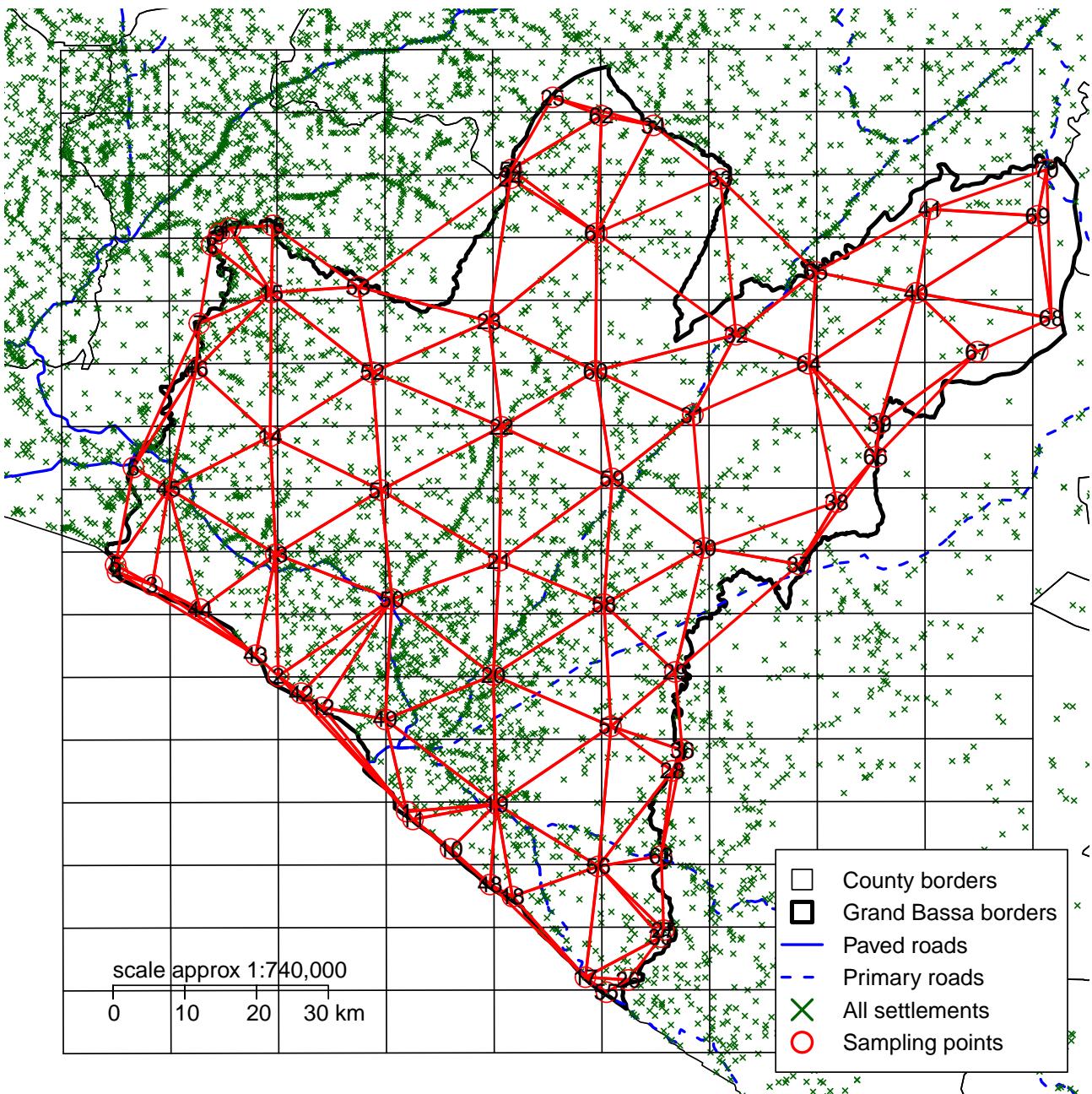


Figure 2.18: Grand Bassa county with a rectangular grid defined by d of 10 km and sampling points moved to the nearest communities showing test triangulation

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.



Figure 2.19: Selecting alternating intersections of the grid in the x and y directions to spread sampling points evenly

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.



Figure 2.20: Montserrado county with a rectangular grid defined by d of 6 km and sampling points moved to the nearest communities showing updated triangulation

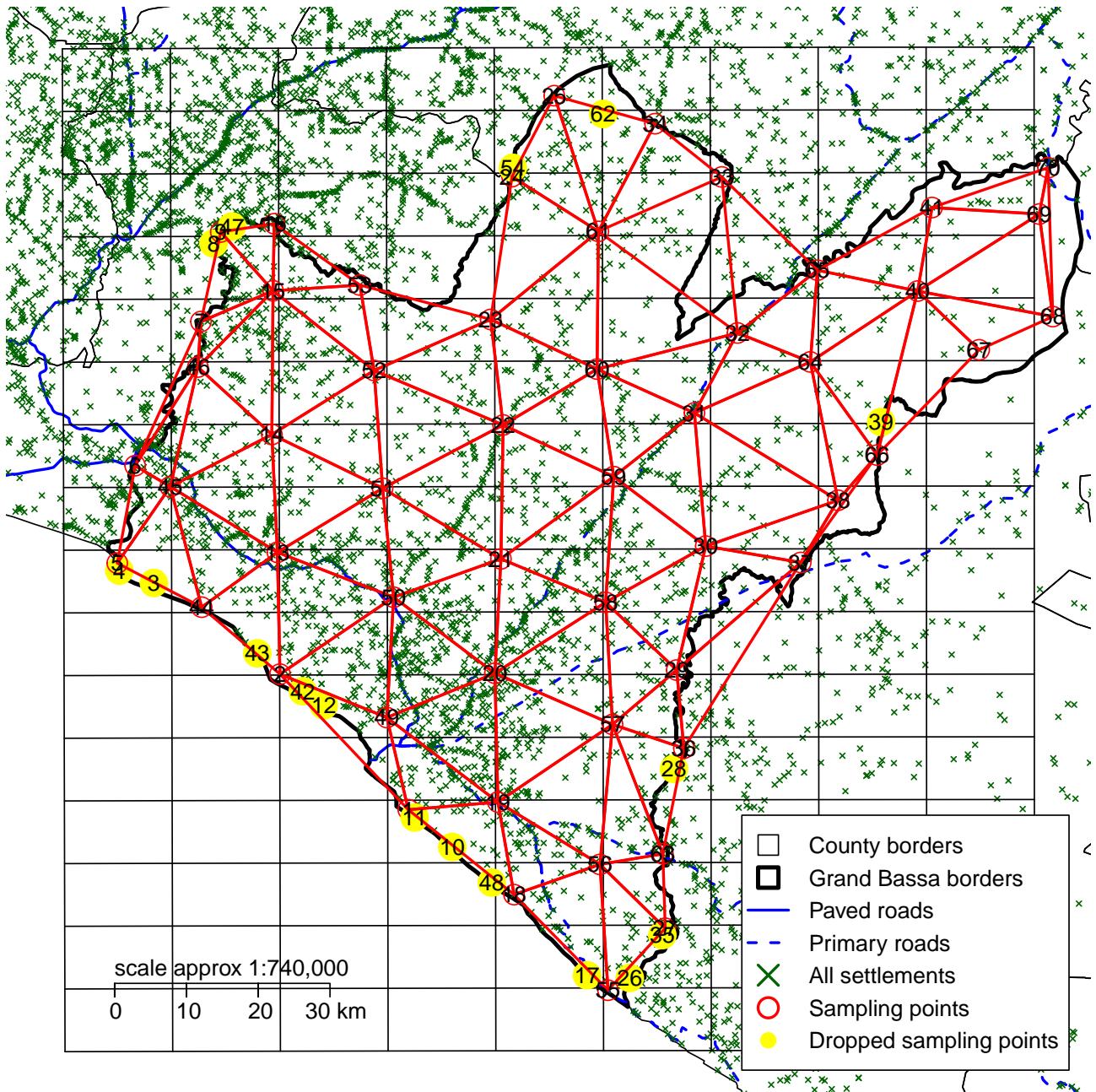


Figure 2.21: Grand Bassa county with a rectangular grid defined by d of 10 km and sampling points moved to the nearest communities showing updated triangulation

Figure 2.20 above show a trial triangulation for Montserrado county with only a few long and narrow triangles. Five sampling points (labelled 2, 25, 32, 33, 39) have been dropped to ensure that there are few long and narrow triangles.

Figure 2.21 above show a trial triangulation for Grand Bassa county with only a few long and narrow triangles. Seventeen sampling points (labelled 3, 4, 8, 10, 11, 12, 17, 26, 28, 35, 39, 42, 43, 47, 48, 54, 62) have been dropped to ensure that there are a 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.

2.2 The second stage sample

The second stage sample is also called *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. For the Liberia S3M, this is the approach that will be used in which there will be 2 concurrent surveys. There will be a survey for CMAM coverage which will be specific for SAM children and a survey for children aged 6 - 59 months and their mothers.

2.2.1 Stage 2 sampling for CMAM coverage survey

For the survey for CMAM coverage, an active and adaptive (snowball) case-finding method can be used in stage 2 sampling to find all or nearly all SAM cases by MUAC in the stage 1 sample ([Myatt et al., 2012](#)). However, given that the Liberia CMAM programme uses both MUAC and weight-for-height as independent criteria for SAM, it is uncertain whether the use of active and adaptive case finding using weight-for-height will find all or nearly all SAM cases

by weight-for-height¹. It is therefore recommended that a full census of children aged 6-59 months in the selected villages in the stage 1 sample be conducted to ensure exhaustivity. This will also be the approach to use in urban areas given that active and adaptive case finding has been known to fail in finding all SAM cases when implemented in urban areas ([Myatt et al., 2012](#)). As an example, the only other known coverage survey conducted in Liberia is that which was conducted by Accion contra la faim in 2011 to assess coverage of CMAM in Greater Monrovia ([Accion Contra la Faim, 2011](#)). In this survey, it was unclear as to whether both MUAC and weight-for-height was used to identify SAM but both active adaptive case finding and mass screening was used to ensure that all or nearly all CAM cases are found. Doing both approaches will be too resource-intensive. A good census sample taken house-to-house or door-to-door should be sufficient.

2.2.2 Stage 2 sampling for children 6-59 months and their mothers

Given that the stage 2 sample for the CMAM coverage survey will utilise a house-to-house or door-to-door of the villages, the stage 2 sample for the survey of children 6-59 months old and their mothers can be done via door/house counting with systematic selection in the first village selected per sampling point.

Generally, in rural villages/communities, the village leader will know how many houses/households are in the village so they can be asked for this information. In urban locations/communities/blocks, this information isn't usually known. So, door counting of the urban location or block will be necessary to know the total number of doors/houses in that particular urban area/location.

Once the number of houses or number of doors has been established, a sampling interval will need to be calculated as follows:

$$n_{\text{Sampling interval}} = \frac{\text{Total number of houses/doors}}{\text{Target sample of children 6-59 months old in village}}$$

The house-to-house/door-to-door sampling to look for SAM children can start and then for every *nth* house/door, the survey for children 6-59 months old will be conducted as well.

Remember that this will be done in the first village selected per sampling point.

¹Capture and re-capture studies done to test the sensitivity of active and adaptive case-finding as a sampling approach to find all or nearly all SAM cases were all implemented using MUAC as the indicator for SAM (see [Wegerdt et al. \(2006\)](#)).

2.2.3 Special considerations for large villages

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 fully enumerated for the SAM coverage survey. If this village is the first village of the sampling point, then the door/house counting with systematic selection will be applied only in the sample of segments selected.

Segmentation involves dividing a community into several parts and taking part of the within-community sample from each **segment**. With simple communities, segmentation is not required and we take a single sample from the entire community using the appropriate sampling method.

2.2.3.1 Segmentation

For more complicated communities we divide the community into several parts or segments, such as a community made up of several clusters:

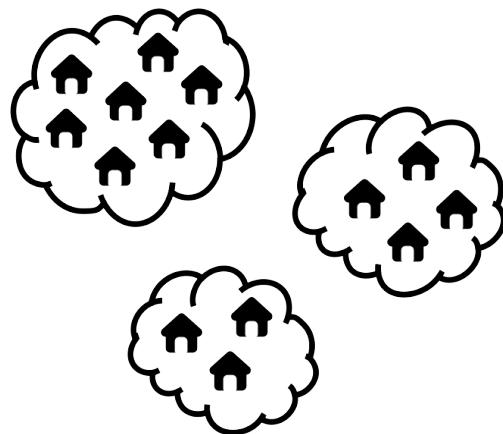


Figure 2.22: Example of a set of clusters of dwellings

or a community made up of several ribbons:



Figure 2.23: Dwellings arranged in several lines

or a mixed community:

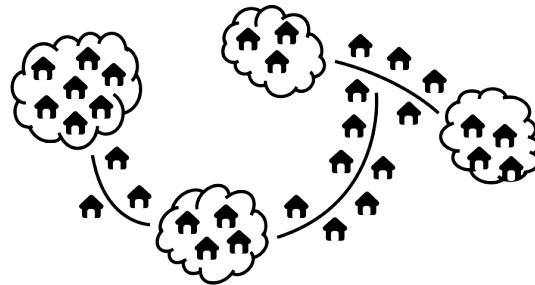


Figure 2.24: Mixture of clusters and ribbons

Segments should be either ribbons or clusters but should **never** contain both a ribbon and a cluster. This is because clusters and ribbons are sampled in different ways.

To select which segments to sample in a very large village or in urban locations/blocks, draw a rough map of the village and number each segment. Then depending on how many segments have been decided to sample (or are able to be sampled), select randomly the numbered segments. The selected segments are then sampled house-to-house/door-to-door. If this village is the first village of the sampling point, then the door/house counting with systematic selection will be applied only in these selected segments.

2.3 Sample size considerations

In general, the sample size needed for proportion-type indicators such as those to reported for this survey can be calculated using the following equation.

$$n = Z^2 \times \frac{p(1 - p)}{c^2}$$

where :

Z = z-value for preferred confidence interval

p = expected indicator proportion/prevalence

c = level of precision

The Z value is usually 1.96 for a 95% confidence interval. The p should usually be based on previous coverage results if available. If not, it is usually appropriate to set p at 50% (0.5) as this results in the highest sample size estimate. The precision (c) for coverage surveys is usually set at $\pm 10\%$ based on standard precision used for immunisation coverage.

Using these values, the typical sample size needed for coverage surveys is about 96.

$$n = 1.96^2 \times \frac{0.5(1 - 0.5)}{0.10^2} \approx 96$$

However, the survey design needs to be taken into account. A cluster survey such as the one that is proposed for the Liberia S3M will need to inflate sample sizes to account for the loss of variance due to the cluster design. This inflation factor is called the *design effect (DEFF)* which is based on the *intraclass correlation coefficient (ICC)*.

Generally, a *DEFF* of 2 is used to multiply the sample size with to account for the loss of variance. This would mean that a sample size of 192 would be the target sample size.

The sample size required will also depend on the indicators being assessed. Following is a further discussion of sample size requirements for the CMAM coverage survey and for the survey for children 6-59 months and their mothers.

2.3.1 Sample size for the CMAM coverage survey

SAM treatment coverage can be demonstrated in the following equation:

$$\text{SAM treatment coverage} = \frac{\text{SAM cases in treatment}}{\text{Total SAM cases}}$$

This indicator requires a sample of under 5 children who have SAM. SAM children are rare (at most 3% of the general population of children under 5). This means that the universe population of SAM children is small hence a finite population correction can be applied to sample size calculations. At the same time, given the active adaptive case finding approach that will be applied to finding SAM cases is known to be exhaustive, *DEFF* would generally be close to 1. Therefore, a sample size of about 96 is generally big enough to estimate SAM treatment coverage with a $\pm 10\%$ precision.

2.3.2 Survey for children aged 6-59 months and their mothers

For vitamin A supplementation, micronutrient powder supplementation, IYCF counselling and ferrous sulphate-folic acid supplementation the sample size of 192 would be the target sample size for each of these indicators.

However, given that this is a compound sample such that the MNP supplementation indicator is only for children 6-23 months old, the 192 sample will need to be inflated for the sample of children 6-59 months old big enough to ensure that at least 192 children aged 6-23 months will be included in the sample. The factor by which to inflate the sample size for children 6-59 months old can be estimated as follows:

$$\frac{\text{Age range in months for 6-59 month old children}}{\text{Age range in months for 6-23 month old children}} = \frac{59 - 6}{23 - 6} = \frac{53}{7} \approx 3$$

The sample size for children aged 6-59 months old will need to be inflated 3 times to ensure that a minimum sample size of 192 children aged 6-23 months will be sampled.

So, our sample size for the survey for children 6-59 months old and their mothers will be 576.

2.4 Urban and rural sample considerations

Currently, the requirements for the survey is to provide coverage estimates for Montserrado rural districts and the whole Grand Bassa county. Urban areas have been specifically excluded.

Given that urban areas would most likely have different characteristics than rural areas, it is more reasonable to treat urban areas as a separate survey area. If urban areas are to be sampled, then urban areas should have its own sample with the same minimum sample size as described above. The implication of this would be an extra survey for each and every urban area to be assessed.

2.5 Overall sampling frame

Given the stage 1 sampling, stage 2 sampling and sample size considerations described above, following are the overall sampling frames for each of two surveys that will be conducted.

2.5.1 Sampling frame for the CMAM coverage survey

Using an S3M approach to stage 1 sampling, thirty (30) stage 1 sampling points will be identified. This is an adaptation of S3M such that instead of first specifying a d value, a minimum number of stage 1 samples is specified based on which the hexagonal grid across each county will be drawn. This would mean that each county will be divided into 30 hexagonal, non-overlapping grids to identify 30 sampling points as the stage 1 sample. If estimates for urban areas will be required, then the urban areas of each county will be separated and a separate grid of 30 hexagons will be drawn on the map of the urban areas.

To determine how many villages near the 30 sampling points to select for the stage 1 sample, it would require estimating how many villages will need to be sampled to obtain the target sample size of 96 SAM children in each county (and urban areas if included). This can be estimated as follows:

$$n_{villages} = \frac{\text{Target sample size}}{n_{\text{average village population all ages}} \times \% \text{ under 5 children} \times \text{SAM prevalence}}$$

The average village population can be estimated using per village population in Grand Bassa county and Rural Montserrado provided by the Liberia Institute of Statistics and Geo-information Services to UNICEF. Using this data, the average village population in Grand Bassa (including urban areas) is 521 and the average village population in Rural Montserrado is 474.

The proportion of children under 5 can be estimated using data from the [US Census Bureau](#) on estimated population for single years. Using this data, the estimated proportion of under five children is about 17%.

The SAM prevalence can be estimated as 3% (based on communication with UNICEF). This is most likely a high estimate considering that this is the estimate for the whole country and not specific to the counties. Also, seasonality of SAM prevalence will need to be taken into account. If the period in which the survey is to be conducted is as season for low prevalence of SAM, this should be considered in the estimation. So, it might be good to use a lower estimate for SAM to ensure there is enough villages sampled to get the target. An estimate of 1.5% (or even lower) is recommended.

Using these values, the number of villages needed to sample in each county (and urban areas if included) can be calculated for each county as follows:

$$\begin{aligned} n_{villages \text{ Grand Bassa}} &= \left[\frac{\text{Target sample size}}{n_{\text{average village population all ages}} \times \% \text{ under 5 children} \times \text{SAM prevalence}} \right] \\ &= \left[\frac{96}{521 \times 0.17 \times 0.015} \right] \\ &= \left[\frac{96}{1.32855} \right] \\ &\approx 73 \end{aligned}$$

$$\begin{aligned}
 n_{\text{villages Rural Montserrado}} &= \left\lceil \frac{\text{Target sample size}}{n_{\text{average village population all ages}} \times \% \text{ under 5 children} \times \text{SAM prevalence}} \right\rceil \\
 &= \left\lceil \frac{96}{474 \times 0.17 \times 0.015} \right\rceil \\
 &= \left\lceil \frac{96}{1.2087} \right\rceil \\
 &\approx 80
 \end{aligned}$$

A total of 73 villages will need to be sampled for Grand Bassa county and a total of 80 villages will need to be sampled for Rural Montserrado.

Given this, it would be ideal to select 3 villages per sampling point as the stage 1 sample for the CMAM survey in Grand Bassa and Rural Montserrado.

Similar calculations will need to be done for each additional urban area to determine how many blocks or enumeration areas will be sampled for the urban area to get the required 96 SAM cases sample size.

2.5.2 Sampling frame for survey of children aged 6-59 months and their mothers

Given a target sample size of 576 and using a stage 1 sample of 30 grids as described above, a single nearest village to each of the 30 sampling points identified in the stage 1 sample will be selected. In each of these villages, the number of children 6-59 months and their mothers can be estimated as follows:

$$\begin{aligned}
 n_{\text{children 6-59 months and mothers}} &= \left\lceil \frac{\text{Target sample size}}{\text{Number of villages}} \right\rceil \\
 &= \frac{576}{30} \approx 20
 \end{aligned}$$

Twenty children 6-59 months and their mothers will need to be sampled from each of the first villages of the 30 sampling points for each county (and urban areas if included).

This will mean that the sampling interval to use for the door/house counting with systematic selection for the stage 2 sampling of the survey for children 6-59 months and their mothers can be calculated as follows:

$$\begin{aligned} n_{\text{Sampling interval}} &= \frac{\text{Total number of houses/doors}}{\text{Target sample of children 6-59 months old in village}} \\ &= \frac{\text{Total number of houses/doors}}{20} \end{aligned}$$

2.6 Estimated time to complete full survey

Given the overall sampling frame for SAM coverage survey and survey for children 6-59 months and their mothers, the estimated time to complete the whole survey will be based on the total number of villages to sample and based on an assumed/estimated rate of completion of a single village by single survey team.

A total of 90 villages will need to be sampled per county (and per urban area if included). Assuming that a single survey team can enumerate one village in a day, then the following lent of survey in days can be considered based on number of survey teams:

Table 2.3: Estimated length of full survey in days given number of survey teams and number of areas to survey

Number of survey teams	Length of survey (per county/area)	Length of survey (2 counties)	Length of survey (2 rural, 1 urban)
5 survey teams	18 days	36 days	54 days
10 survey teams	10 days	20 days	30 days
15 survey teams	6 days	12 days	18 days

Chapter 3

Indicators

3.1 CMAM coverage

CMAM coverage usually pertains to coverage of SAM treatment. Historically, there have been two coverage estimators in common use: **point** and **period** coverage.

Point coverage is the number of current SAM cases in a treatment programme divided by the total number of current SAM cases.

Point coverage uses data for current cases only. It is calculated using the following formula:

$$\text{Point coverage} = \frac{C_{in}}{C_{in} + C_{out}}$$

where :

C_{in} = current SAM cases in the programme

C_{out} = current SAM cases out of the programme

Point coverage provides a snapshot of programme performance, putting a strong emphasis on the effectiveness and timeliness of case-finding and recruitment ([Myatt et al., 2012](#)).

Period coverage, on the other hand, uses data for both current and recovering cases. It is calculated using the following formula:

$$\text{Period coverage} = \frac{C_{in} + R_{in}}{C_{in} + C_{out} + R_{in}}$$

where :

R_{in} = recovering SAM cases in the programme

Period coverage is the number of current and recovering cases in a treatment programme divided by all current SAM cases and recovering cases. It approximates treatment coverage much better (albeit with limitations) as it accounts for children who are no longer cases but are in the programme.

Point and period coverage both have their limitations.

Point coverage of a programme with good case-finding and recruitment and short lengths of stay can be misleadingly low because there are too few current cases. For example, a coverage survey found:

Table 3.1: Scenario 1 - coverage survey results from a programme with good case-finding and recruitment and short lengths of stay

Number of current cases	2
Number of current cases in the programme	0
Number of current cases not in the programme	2
Number of recovering cases in the programme	34

In this scenario, the point coverage estimator returns:

$$\text{Point coverage} = \frac{C_{in}}{C_{out}} = \frac{0}{2} = 0 = 0\%$$

but the period estimator returns:

$$\text{Period coverage} = \frac{0 + 34}{0 + 34 + 2} = 0.944 = 94.4\%$$

In this regard, the point coverage estimate penalises good performance and the period coverage most likely better depicts the coverage of the programme.

On the other hand, a programme with poor case-finding and recruitment and long lengths of stay due to late presentation and/or late admission may have a period coverage that is misleadingly high because of high number of recovering cases. In such a scenario, the two estimators will yield very different results. For example:

Table 3.2: Scenario 2 - coverage survey results from a programme with good case-finding and recruitment and short lengths of stay

Number of current cases	12
Number of current cases in the programme	3
Number of current cases not in the programme	9
Number of recovering cases in the programme	22

$$\text{Point coverage} = \frac{C_{in}}{C_{out}} = \frac{3}{12} = 0.250 = 25.0\%$$

but the period estimator returns:

$$\text{Period coverage} = \frac{3 + 22}{3 + 22 + 9} = 0.735 = 73.5\%$$

In this example, point coverage is the more reflective coverage of the programme.

It should be noted also that period coverage has a tendency to an overestimation bias. This is because the current period coverage estimator does not take into account cases of acute malnutrition who have recovered spontaneously but were never enrolled in any treatment programme.

An estimator of coverage that does include both recovering cases that are in the programme and recovering cases that are not in the programme and, thus, provides an unbiased estimator of overall programme performance is:

$$\text{Single coverage} = \frac{C_{in} + R_{in}}{C_{in} + R_{in} + C_{out} + R_{out}}$$

where :

R_{out} = Recovering SAM cases not in the programme

It is for these reasons that the coverage assessment technical guide ([Myatt et al., 2012](#)) recommends that only one of these estimators be reported and the choice of estimator to report should be guided by specific programme features and characteristics (such as lengths of stay in the programme) that would justify the choice of reported estimator. In a recent Epicentre review ([Epicentre, 2015](#)) this has been highlighted as a source of confusion and issues given the possibility of period coverage being chosen more as a coverage estimator rather than point coverage because of it being a higher estimate even if the programme characteristics do not merit its use. The review suggests that both coverage indicators could be reported, with sufficient context (e.g. on length of stay, timeliness of admissions, etc.) to allow for their interpretation.

In response to the review and to address the limitations of the coverage estimators, development work has been conducted on further developing and improving the coverage estimators to address this confusion and the issues around them ([Balegamire et al., 2015](#)). This work focused primarily on improving the period coverage estimator to address the overestimation bias described earlier and to make it more closely approximate the treatment coverage estimator formula shown above.

The problem with this estimator is that R_{out} (i.e. the number of recovering cases that are not in the programme) is unknown and may be difficult to collect accurately. This problem of estimating the number of recovering cases not in the programme (R_{out}) may be addressed using a simple mathematical model¹ proposed by [Balegamire et al. \(2015\)](#).

¹This model assumes that incidence of acute malnutrition and programme coverage do not vary rapidly over time.

$$\frac{C_{in} + C_{out}}{C_{in}} \approx \frac{R_{in} + k \times R_{out}}{R_{in}}$$

where :

k = a correction factor

R_{out} can then be expressed in terms of the known variables:

$$R_{out} \approx \left\lfloor \frac{1}{k} \times \left(R_{in} \times \frac{C_{in} + C_{out}}{C_{in}} - R_{in} \right) \right\rfloor$$

Given the possibility that no cases in the programme (C_{in}) are found, the calculation is adjusted by adding 1 to C_{in} . To arrive at a whole number value for R_{out} , the calculation is rounded off towards zero.

$$R_{out} \approx \left\lfloor \frac{1}{k} \times \left(R_{in} \times \frac{C_{in} + 1 + C_{out}}{C_{in} + 1} - R_{in} \right) \right\rfloor$$

This leaves the problem of deciding a suitable value for the correction factor (k). A reasonable candidate for k is the ratio of the mean length of an untreated episode to the mean length of a CMAM treatment episode.

$$k = \frac{\text{Mean length of untreated episode}}{\text{Mean length of a treatment episode}}$$

Possible value for mean length of untreated episode is 7.5 months (Garenne et al., 2009) which is the common value used when estimating programme case-loads from prevalence estimates (Myatt, 2012). Mean length of a treatment episode can be estimated by calculating the mean length of stay in the CMAM programme using routine monitoring data. In general, A value of

2.5 months could be used in the absence of better information or when the validity of routine programme monitoring data is suspect. Using these values, k is:

$$k = \frac{\text{Mean length of untreated episode}}{\text{Mean length of a treatment episode}} = \frac{7.5}{2.5} = 3$$

The inclusion of recovering cases means that the single coverage estimate is mathematically constrained to return a coverage estimate that is greater than or equal to the point coverage estimate. The underestimation present in the point coverage estimate has, to some extent, been corrected. The inclusion of recovering cases that are not in the programme means that the single coverage estimator is mathematically constrained to return a coverage estimate that is less than or equal to the period coverage estimate. The overestimation present in the period coverage estimate has, to some extent, been corrected.

Given this single coverage estimator, a shift in terminology is proposed that is more descriptive and specific with regard to what the estimator is actually measuring, allowing both measures to be reported together without confusion. **Point coverage** is now named *case-finding effectiveness* to more precisely reflect it as a measure of the programme's ability to find and recruit current cases. This indicator assesses how good the treatment programme is in finding cases of SAM and then getting them to treatment. **Period coverage** that has been improved into the single coverage metric is now named *treatment coverage* as this is the estimator that approximates this coverage indicator the closest.

3.1.1 Requirements

This indicator will require additional programmatic information to be operationalised specific for Liberia.

It would be ideal that a Liberia-specific value for mean length of a treatment episode be estimated using routine programme data from a selection of cured cases from a selection of health centres providing CMAM services in the two counties in which the survey is to be implemented. This data should be collected from the beneficiary cards of cured cases. For each county, a total of 30 health centres from the full list of health centres in the county can be selected systematically. Of these 30 health centres, a sample of 30 records of discharged cured cases should be selected. From these records, data on date of admission and date of discharge should be extracted collated into a data format (ideally a comma-separated value or CSV file) with each of these values in a separate column.

This data will allow the calculation of the mean length of stay in the CMAM programme of discharge cured cases.

$$\text{Mean length of stay} = \frac{\text{Length of stay in programme (days)}}{\text{Number of records}}$$

For data collection, enumerators will require:

1. A height board;
2. A digital weighing scale; and,
3. A MUAC tape.

Traditionally, if data collection is done using pen and paper, enumerators are provided with z-score tables for weight-for-height so that they can determine if child is SAM. Since this survey will utilise digital data collection (see Section @ref(#analysis)), the digital form will determine whether the child is SAM by MUAC and/or weight-for-height and/or oedema and will instruct the enumerators on what to do next if the child is a SAM case.

Referral slips will also be needed so that children who are identified as SAM but not in the programme can be referred.

3.2 Vitamin A supplementation

The standard estimator for vitamin A supplementation is the proportion of children aged 6-59 months who received two age-appropriate doses of vitamin A in the past 12 months.

In standard surveys such as the DHS and MICS, this indicator is adjusted to a recall of 6 months for a single age-appropriate dose of vitamin A.

In order to determine whether supplementation with vitamin A is age appropriate, vitamin A supplementation should first be assessed on the child's health card. Provision of vitamin A card is usually recorded on the child's health card with the corresponding does given. If health card not available or if health card is lost or if the child doesn't have a health card at all, then the mother/caregiver will have to be asked whether their child has received vitamin A in the past 6 months and respond by recall. If the mother/caregiver says yes, then the next question to ask will be which type of gel capsule was provided. The blue vitamin A gel capsule containing 100,000 IU of vitamin A is given to children 6-11 months. The red vitamin A gel capsule containing 200,000 IU of vitamin A is given to children 12 - 59 months. A sample of the blue and the red gel capsule (or a photo of the capsules) can be used to aid the mother/caregiver in answering this question.

Given this, two indicators can be reported on vitamin A supplementation.

1. Any vitamin A supplementation in the past 6 months.
2. Age-appropriate vitamin A supplementation in the past 6 months.

3.2.1 Requirements

Either samples and/or photos of both the blue and red gel vitamin A capsules will be needed per enumerator to use in aiding the mother to answer which type of vitamin A gel capsule has been provided to the child.

3.3 Iron-folic acid (IFA) supplementation for pregnant women

Population-based surveys typically report the percentage of women with a live birth in the two to five years before the survey who received and took IFA supplementation during their most recent pregnancy. Because antenatal care (ANC) is typically the main platform for IFA supplement distribution for pregnant women, survey questions on antenatal care attendance and timing of the first antenatal care visit can provide information on the use of this platform to deliver IFA supplementation. [Sununtnasuk et al. \(2015\)](#) propose a falter point framework² that utilises four indicators that proxy the four critical points at which the ANC approach to IFA distribution might falter in IFA supplementation coverage to pregnant women. These indicators are:

1. At least one ANC visit during most recent pregnancy
2. Receipt or purchase of IFA tablet/s
3. IFA consumption
4. Adherence to 180 days of supplementation

3.3.1 Requirements

It would be ideal for each enumerator to have a sample/s and/or photos of an IFA tablet to show the respondent as an recall aid or prompt. Given that some mothers may have to recall as far back as 5 years and that they might not be able to remember that the tablet is called, showing them a sample of the tablet might help them remember if they have taken it.

²Similar to a bottleneck framework and consistent with [Tanahashi \(1978\)](#) hierarchical model of coverage.

3.4 Micronutrient powder supplementation

The indicator for coverage of micronutrient powder supplementation is the proportion of children aged 6-23 months who consume micronutrient powder supplements. Depending on the programme protocol on mechanism of distribution and effective intake of MNP, a full indicator set on MNP supplementation can be devised that will be similar to the IFA supplementation falter point or bottleneck framework. For example, if MNPs were being provided through the health centres or health posts, then the following indicators can be assessed hierarchically:

1. Health centre / health post attendance in the past month
2. Awareness of MNP
3. Consumption of MNP

3.4.1 Requirement

It would be ideal to provide enumerators with sample/s and/or photo/s of MNP sachets used in Liberia. If there are various brands of MNP distributed, it would be good to have sample/s and/or photo/s of each one. These sample/s and/or photo/s can be used to help mothers or caregivers recall or know what an MNP is and be able to ascertain whether they do know it or have heard of it and whether their child has consumed it.

3.5 IYCF counselling

There are no standard indicators for IYCF counselling. Any indicator developed for this programme will depend on the mechanics of how the IYCF counselling is delivered and who the target beneficiaries are. In terms of mechanism, what is known so far is that these sessions are delivered via the health clinic/health post and that the target beneficiaries are pregnant women. Given this, similar approach to the IFA supplementation coverage of falter points/bottle necks can be used with the following indicators:

1. At least one ANC visit during most recent pregnancy
2. Awareness of IYCF counselling (have they been advised IYCF counselling when they attended ANC)
3. Attendance to IYCF counselling

3.5.1 Requirement

It would be ideal to have the training materials used for IYCF counselling and the materials used when conducting IYCF counselling. These materials can be used as a recall aid for mothers/caregivers who are being asked about whether they know or have heard about IYCF counselling and whether they have attended.

Chapter 4

Questionnaire

The following are sample/template questionnaires used for the two types of surveys that will be implemented.

4.1 CMAM coverage survey instruments

CMAM coverage surveys use primarily use two forms. The first form is used to collect coverage data from SAM children found during the survey. Given that this survey will use house-to-house/door-to-door sampling for stage 2, then it would be necessary to record all data from all children that are measured with MUAC, weight-for-height and oedema. The following tabular form can be used for this purpose:

SAM Coverage Data Collection Form

Figure 4.1: SAM coverage survey sample/template form

The data collected using the tabular forms will allow estimation of coverage. They do not, however, allow you to know the reasons for coverage failure. To collect this data we apply a “barriers” questionnaire to the mothers/carers of uncovered SAM cases. Here is an example of a barriers questionnaire:

SAM Coverage Barriers Data Collection Form

PSU ID	ID
__ __	__ __

B1: Do you know that your child is malnourished or thinner than he/she should be? YES NO**B2:** Do you know that there is a programme that can treat malnourished children? YES NOIf **NO**, thank mother/carer and terminate interview.**B3:** What do you call this programme? _____**B4:** Where is this programme? _____**B5:** Why is this child not attending this programme?*Do not prompt. Probe "Any other reasons?"*

- Programme site is too far away
- No time/too busy to attend the programme
- Carer/mother cannot travel with more than one child
- Woman cannot travel alone
- Carer is ashamed to attend programme
- Difficulty with childcare
- Child has been rejected by the programme

Record all other reasons given:

B6: Has this child ever been to the program site or examined by program staff? YES NOIf **NO**, thank mother/carer and terminate interview.**B7:** Why is this child not in the programme now?

- Previously rejected
- Defaulted
- Discharged as cured
- Discharged as not cured

Thank mother/carer. Issue a referral slip. Inform mother/carer of site and date to attend.

Figure 4.2: SAM coverage barriers survey sample/template form

4.2 Survey for children 6-59 months and their mothers

For the survey for children 6-59 months, following is a sample/template questionnaire.

Children 6-59 months Data Collection Form		
PSU ID	County	Village Name
<input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> 1 = Grand Bassa 2 = Rural Montserrado	<input type="text"/>
Administrative Data		
ad1	Date of data collection	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
ad2	Team number	<input type="text"/> <input type="text"/> <input type="text"/>
ad3	Name of enumerator	<input type="text"/>
ad4	Consent?	1 = YES 2 = NO
<i>If consent is not given then stop the interview and move to next respondent.</i>		
Identifying Data		
psu	PSU (cluster) number	<input type="text"/> <input type="text"/> <input type="text"/>
hh	Household identifier	<input type="text"/> <input type="text"/> <input type="text"/>
Mother Information		
m1	Mother's identifier	<input type="text"/> <input type="text"/> <input type="text"/>
m2	Mother's age (in full years)	<input type="text"/> <input type="text"/> <input type="text"/>
m3	How many children 6-59 months do you have?	<input type="text"/> <input type="text"/> <input type="text"/>
Iron-Folic Acid (IFA) Supplementation		
ifa1	During your pregnancy for your youngest child, did you see anyone for antenatal care?	1 = YES 2 = NO 99 = Don't know/not sure <input type="text"/> <input type="text"/> <input type="text"/>
ifa2	During your pregnancy for your youngest child, did you receive any information about taking iron-folic acid tablets (show sample/photo)?	1 = YES 2 = NO 99 = Don't know/not sure <input type="text"/> <input type="text"/> <input type="text"/>
ifa3	During your pregnancy for your youngest child, did you receive and/or purchase iron-folic acid tablets (show sample/photo)?	1 = YES 2 = NO 99 = Don't know/not sure <input type="text"/> <input type="text"/> <input type="text"/>
	<i>If NO, go to ifa3a; If DON'T KNOW/NOT SURE, skip to icf1; If YES, go to ifa4</i>	
ifa3a	Why did you not receive or purchase iron-folic acid tablets? <i>Do not prompt. Probe "Any other reasons?"</i>	<input type="checkbox"/> Health centre ran out <input type="checkbox"/> Took too long to get tablets <input type="checkbox"/> Too expensive Other reasons: <input type="text"/> <input type="text"/>

Figure 4.3: Children 6-59 months old and their mothers survey sample/template form

ifa4	During your pregnancy for your youngest child, did you take iron-folic acid tablets (show sample/photo)?	1 = YES 2 = NO 99 = Don't know/not sure	__ __
<i>If NO, go to ifa4a; If DON'T KNOW/NOT SURE, skip to icf1; If YES, go to ifa5</i>			
ifa4a	Why did you not take iron-folic acid tablets? <i>Do not prompt. Probe "Any other reasons?"</i>	__ Concerns about side effects __ Don't need it __ I was told not to take it __ I don't think it helps Other reasons: _____ _____	__ __ __
ifa5	During the whole duration of your pregnancy for your youngest child, how many days did you take the iron-folic acid tablets (show sample/photo)?	__ __ __	

IYCF Counselling			
icf1	During your pregnancy for your youngest child, did you receive any information or hear about any service that provided counselling for pregnant women on infant and young child feeding?	1 = YES 2 = NO 99 = Don't know/not sure	__ __
icf2	During your pregnancy for your youngest child, did you attend any counselling session for pregnant women on infant and young child feeding?	1 = YES 2 = NO 99 = Don't know/not sure	__ __
<i>If NO, go to icf2a; If DON'T KNOW/NOT SURE, skip to ; If YES, go to icf3.</i>			
icf2a	Why did you not attend the counselling sessions? <i>Do not prompt. Probe "Any other reasons?"</i>	__ Timing was not convenient __ Not interested __ Do not trust counsellor __ Don't think I need it Other reasons: _____ _____	__ __
icf3	How many counselling sessions did you attend?	__ __	

Figure 4.4: Children 6-59 months old and their mothers survey sample/template form

Child Information			
	<i>If m3 > 1, randomly select one child from the mother's children.</i>		
c1	Age in months of respondent child	<input type="text"/> <input type="text"/>	
	<i>If c1 ≥ 23, go to mnp1; If c2 > 23, go to vit1</i>		
Micronutrient Powder (MNP) Supplementation			
mnp1	Have you heard about [LOCAL NAME OF MNP] or seen [show sample/photo of MNP]?	1 = YES 2 = NO 99 = Don't know/not sure	<input type="text"/> <input type="text"/>
mnp2	Have you received or bought [LOCAL NAME of MNP] (show sample/photo of MNP) for your child aged 6-23 months?	1 = YES 2 = NO 99 = Don't know/not sure	<input type="text"/> <input type="text"/>
	<i>If NO, go to mnp2a; If DON'T KNOW/NOT SURE, go to vit1; IF YES, go to mnp3</i>		
mnp2a	Why did you not receive/buy [LOCAL NAME of MNP] (show sample/photo of MNP) for your child aged 6-23 months? <i>Do not prompt. Probe "Any other reasons?"</i>	<input type="checkbox"/> Health centre ran out <input type="checkbox"/> Took too long to get MNP <input type="checkbox"/> Too expensive <input type="checkbox"/> Child doesn't need it <input type="checkbox"/> Heard it doesn't work/help Other reasons: <hr/> <hr/>	
mnp3	Did you give your child aged 6-23 months the [LOCAL NAME OF MNP] to eat?	1 = YES 2 = NO 99 = Don't know/not sure	<input type="text"/> <input type="text"/>
	<i>If NO, go to mnp3a; If DON'T KNOW/NOT SURE, go to vit1; IF YES, go to mnp4</i>		
mnp3a	Why did you not give the [LOCAL NAME OF MNP] to your child aged 6-23 months to eat? <i>Do not prompt. Probe "Any other reasons?"</i>	<input type="checkbox"/> Too expensive <input type="checkbox"/> Not available in the market <input type="checkbox"/> Do not need MNP <input type="checkbox"/> Heard other's bad experience <input type="checkbox"/> Advised not to use it <input type="checkbox"/> Not seen other mothers use it <input type="checkbox"/> Don't trust the product <input type="checkbox"/> Using another product <input type="checkbox"/> Haven't seen it Other reasons: <hr/> <hr/>	
mnp4	In the last week, how many times did your child aged 6-23 months old eat [LOCAL NAME OF MNP]?	<input type="text"/> <input type="text"/>	

Figure 4.5: Children 6-59 months old and their mothers survey sample/template form

Vitamin A Supplementation			
vit1	In the past 6 months, has your child received drops from a capsule like this one (show vitamin A sample or photo)?	1 = YES 2 = NO 99 = Don't know/not sure	<input type="checkbox"/> <input type="checkbox"/>
	<i>If NO, go to vit1a; If DON'T KNOW/NOT SURE, thank mother and end interview; If YES, go to vit2.</i>		
vit1a	Why did your child not receive (show vitamin A capsule sample or photo)? <i>Do not prompt. Probe "Any other reasons?"</i>	<input type="checkbox"/> Health centre ran out <input type="checkbox"/> Took too long to get drops <input type="checkbox"/> Child doesn't need it <input type="checkbox"/> Heard it doesn't work/help Other reasons: <hr/> <hr/> <hr/>	
vit2	What type of capsule (show sample or photo of blue and red capsule) did your child get drops from?	1 = blue capsule 2 = red capsule 99 = Don't know/not sure	<input type="checkbox"/> <input type="checkbox"/>
vit3	How many months ago did your child receive drops from (show vitamin A sample or photo)?	<input type="checkbox"/>	

Figure 4.6: Children 6-59 months old and their mothers survey sample/template form

Chapter 5

Data management and analysis

Data will be collected using an electronic data entry system based on the **Open Data Kit (ODK)** standard that runs on the Android® operating software (OS) platform for mobile devices. The study instrument (see Section 4) will be encoded into the electronic data entry system platform and will be served out of either a local computer server or a remote/cloud-based server¹. Each study team will be provided with mobile devices running on Android® OS that have been configured with an application that will receive the electronic data form. All measurements and answers by respondents will then be recorded on the mobile devices and will then be transmitted to the local server or the remote server whenever there is a mobile phone and / or WiFi signal.

Appropriate data check mechanisms will be put in place using the data check systems available with the ODK system. Spot checks by study supervisor/s will be done in the field to ensure that enumerators are performing the measurements correctly, administering the study instrument correctly and entering all data accordingly. Further checks will be performed by the survey manager on the data as needed as they are received by the local server or remote server.

The digital data collection forms based on the Open Data Kit (ODK) standard using XLS-form and XML is available at <https://github.com/validmeasures/liberiaS3Mforms>. The form faithfully replicates the paper-based forms described in Section 4. The form also faithfully replicates the sampling design described by the survey design such that proposed sampling interval within a primary sampling unit (PSU) is implemented as the survey team goes house-to-house finding SAM cases. The modules on IFA supplementation, IYCF counselling, vitamin A supplementation and MNP supplementation coverage appear appropriately on the digital forms when the correct household based on the sampling interval is being interviewed. The forms also appear appropriately based on eligibility of the child/mother as respnodents to the specific survey modules.

These forms can be viewed and tested using an Android-based mobile device. Instructions for

¹This will depend on what resources are available to the Liberia Institute of Statistics and Geo-Information Services. Valid International has resources to implement either a local server or a remote server solution.

usage can be found at <https://github.com/validmeasures/liberiaS3Mforms>.

Data analysis will be done using R language for statistical computing ([R Core Team, 2018](#)) through bespoke analytic scripts put together into application-based workflow which can be used by non-R users. The application that will be created for this study will be designed such that an automatic reporting format is produced including maps, figures and tables of results.

5.1 Analytical approach for estimating coverage indicators

It is important to note that data analysis procedures need to account for the sample design. All major statistical analysis software can do this (details vary). There are two things to note:

- This survey is a two-stage sample. Subjects are sampled from a small number of primary sampling units (PSUs).
- This survey is **not** prior weighted. This means that per-PSU sampling weights will be needed. These are usually the populations of the PSU.

This sample design will need to be specified to your statistical analysis software. Failing this, the analysis may produce estimates that place undue weight to observations from smaller communities with confidence intervals with lower than nominal coverage (i.e. they will be too narrow).

For this survey, the *blocked weighted bootstrap* estimation approach will be used:

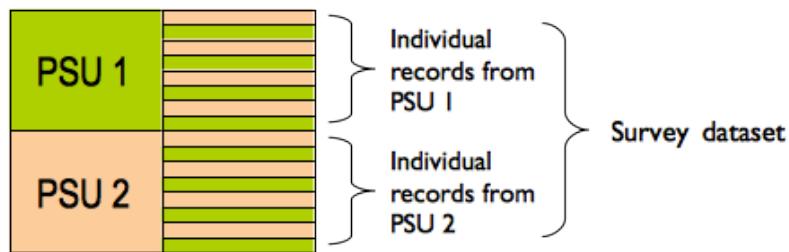
- **Blocked** : The block corresponds to the PSU or cluster.
- **Weighted** : The sampling procedure for this survey does not use population proportional sampling to weight the sample prior to data collection as is done with SMART type surveys. This means that a posterior weighting procedure is required. The “roulette wheel” algorithm to weight (i.e. by population) the selection probability of PSUs in bootstrap replicates will be utilised.

A total of m PSUs are sampled *with-replacement* from the survey dataset where m is the number of PSUs in the survey sample. Individual records within each PSU are then sampled *with-replacement*. A total of n' records are sampled *with-replacement* from each of the selected PSUs where n is the number of individual records in a selected PSU. The resulting collection of records replicates the original survey in terms of both sample design and sample size. A large number of replicate surveys are taken (minimum of $r = 399$ replicate surveys but this can be changed). The required statistic (e.g. the mean of an indicator value) is applied to each replicate survey. The reported estimate consists of the 50th (point estimate), 2.5th (lower 95% confidence limit), and the 97.5th (upper 95% confidence limit) percentiles of the distribution of the statistic observed across all replicate surveys. The blocked weighted bootstrap procedure is outlined in Figure 5.1.

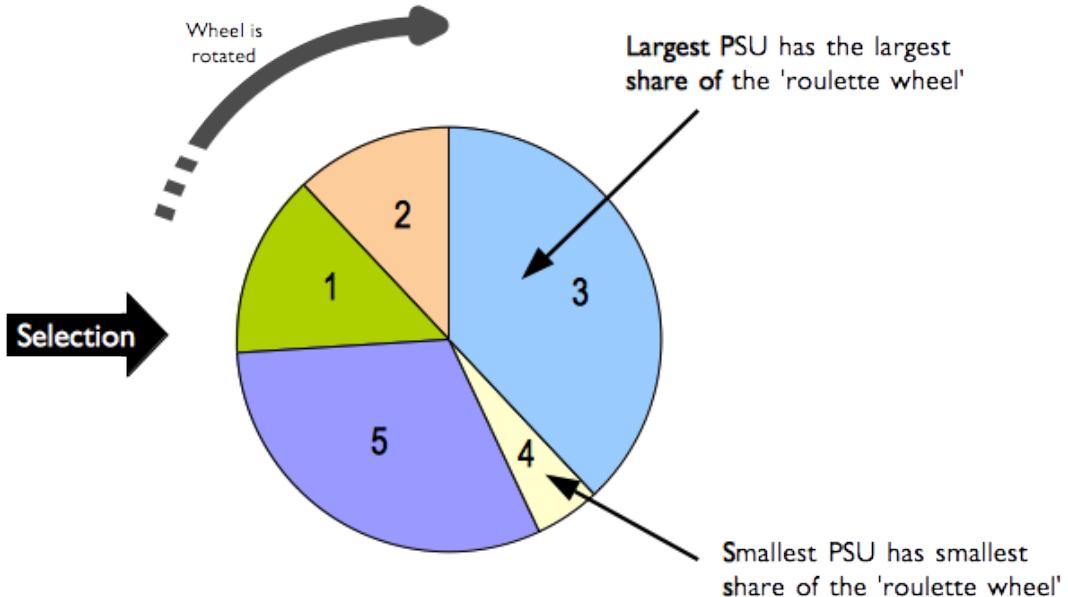
The principal advantages of using a bootstrap estimator are:

- Bootstrap estimators work well with small sample sizes.
- The method is *non-parametric* and uses empirical rather than theoretical distributions. There are no assumptions of things like normality to worry about.
- The method allows estimation of the sampling distribution of almost any statistic using only simple computational methods.

Data are collected using a two-stage cluster design:



PSUs are selected from the survey dataset *with-replacement* and with probability proportional to population size (PPS) using a *roulette wheel* algorithm :



PSUs are selected *with-replacement* and proportional to population size (PPS) :



Individual records are selected *with-replacement* from within each PSU to create a *replicate survey*. The estimator is applied to the replicate survey and the result recorded. This process is repeated many times. The estimate of the indicator value is made from the distribution of the results from each replicate survey which is the *empirical sampling distribution* of the indicator :

Estimate and 95% confidence interval from 50th, 2.5th, and 97.5th percentiles of the empirical (i.e. observed) sampling distribution of the indicator

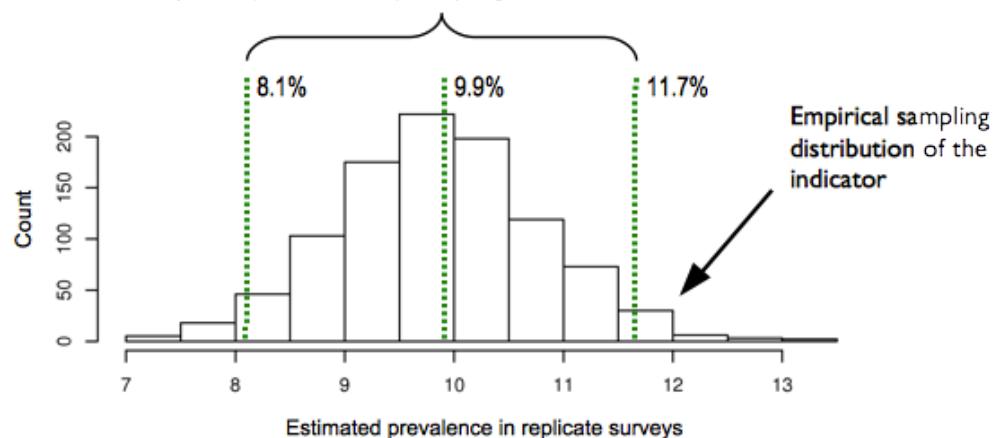


Figure 5.1: The blocked weighted bootstrap

5.2 Analytical approach for mapping coverage indicators

The indicator mapping will create a surface map of indicator values using spatial interpolation. There are various approaches and methods of spatial interpolation, the main differences are determined by the weights applied to the point dataset to estimate values at each of the unknown points of the surface map. For the Liberia coverage survey, spatial interpolation will be performed using the inverse distance weighting (IDW) method. As the name implies, the IDW method uses weights that are inversely proportional to the distance of a point being estimated from the sampling point locations ([Isaaks and Srivastava, 1989](#); [Diggle and Ribeiro, 2007](#); [Diggle, 2013](#)). This can be mathematically demonstrated as follows:

$$\hat{v} = \frac{\sum_{i=1}^n \frac{1}{d_i^p} v_i}{\sum_{i=1}^n \frac{1}{d_i^p}}$$

where :

$d_1 \dots d_n$ = distances from each n sampling points to estimation point

p = power of the distance

$v_1 \dots v_n$ = sample values

The power of the distance p is an important aspect of the IDW method for point estimation. The influence of p to the weights applied to the point estimation is such that as p approaches 0, the weights become more similar, thereby giving more weight to the nearest sample values. As p approaches ∞ , the weights become more different from each other, thereby giving more weight to the closest sample. The power of the distance p has been traditionally set at 2 for convenience and ease of calculations. In theory, given a set p , IDW calculations can be performed using manual calculations aided by a spreadsheet and / or a calculator as it requires fewer calculations. For the Liberia Coverage Survey, p will be initially set at 2 and then cross-validation (see below) will be applied to optimise p to a value that minimises the estimation errors at each of the sampling point locations.

Cross-validation is a technique applied to validate predictive models. It assesses how accurately the predictive model performs in practice. IDW is one of the simplest model-based interpolation methods available, but ideally would still require a form of cross-validation to determine the optimal value of the distance power p (described above).

A two-fold cross validation ([Bivand et al., 2013](#)) in which data points are randomly split into

two sets of equal size, with one set assigned as the validation data for testing the model, and the other set as the training data. The validation data is then interpolated using the IDW method with an initial p of 2 and the resulting predictions were compared with the training data. Comparison is made using the sum of the squared residuals between the predicted values and the observed values to report errors. Optimisation is then performed by replicating the two-fold cross validation process 100 times using randomly generated values for p . Out of these replicates, the value of p that provided prediction results with the minimum errors is selected as the distance power for the eventual interpolation performed.

Bibliography

- Accion Contra la Faim (2011). Semi-Quantitative Evaluation of Access & Coverage (SQUEAC) Greater Monrovia Liberia. Technical report, Accion Contra la Faim, London.
- Balegamire, S., Siling, K., Alvarez, J. L., Guevarra, E., Woodhead, S., Norris, A., Fieschi, L., Binns, P., and Myatt, M. (2015). A single coverage estimator for use in SQUEAC, SLEAC, and other CMAM coverage assessments. *Field Exchange*.
- Bivand, R., Pebesma, E., and Gómez-Rubio, V. (2013). *Applied Spatial Data Analysis with R*. Use R! Springer New York.
- Diggle, P. (2013). *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns, Third Edition*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis.
- Diggle, P. and Ribeiro, P. (2007). *Model-based Geostatistics*. Springer Series in Statistics. Springer.
- Epicentre (2015). Open review of coverage methodologies: Questions, comments and ways forward. Technical report, Epicentre.
- Garenne, M., Willie, D., Maire, B., Fontaine, O., Eeckels, R., Briand, A., and Van den Broeck, J. (2009). Incidence and duration of severe wasting in two African populations. *Public Health Nutrition*, 12(11):1974–1982.
- Isaaks, E. and Srivastava, R. (1989). *Applied Geostatistics*. Oxford University Press.
- Myatt, M. (2012). How do we estimate case load for SAM and / or MAM in children 6 – 59 months in a given time period ? CMAM Forum.
- Myatt, M., Guevarra, E., Fieschi, L., Norris, A., Guerrero, S., Schofield, L., Jones, D., Emru, E., and Sadler, K. (2012). *Semi-Quantitative Evaluation of Access and Coverage (SQUEAC)/ Simplified Lot Quality Assurance Sampling Evaluation of Access and Coverage (SLEAC) Technical Reference*. FHI 360/FANTA, Washington, DC.
- R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

- Sununtnasuk, C., D'Agostino, A., and Fiedler, J. L. (2015). Iron+folic acid distribution and consumption through antenatal care: identifying barriers across countries. *Public Health Nutrition*, 19(04):732–742.
- Tanahashi, T. (1978). Health service coverage and its evaluation. *Bulletin of the World Health Organization*, 56(2):295–303.
- Wegerdt, J., Zanchettin, M., and Myatt, M. (2006). Assessing sensitivity of case-finding using capture-recapture techniques. *Field Exchange*, (27):2.