

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

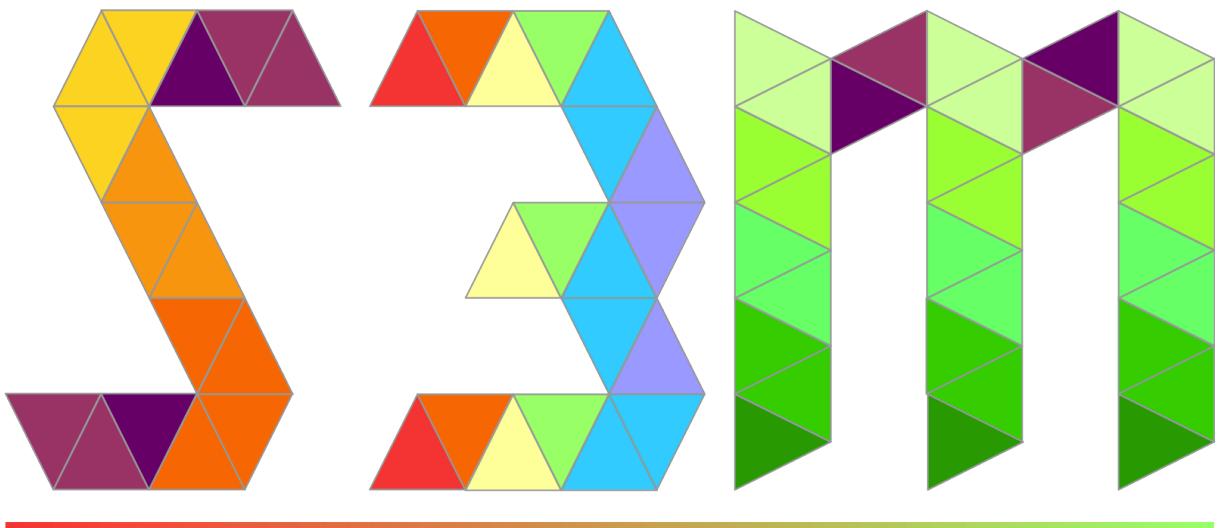
Valid International

2018-08-01

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	37
2.4 Urban and rural sample considerations	38
3 Indicators	41
3.1 CMAM coverage	41
4 Questionnaire	45
5 Analysis	47

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 sulfate-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

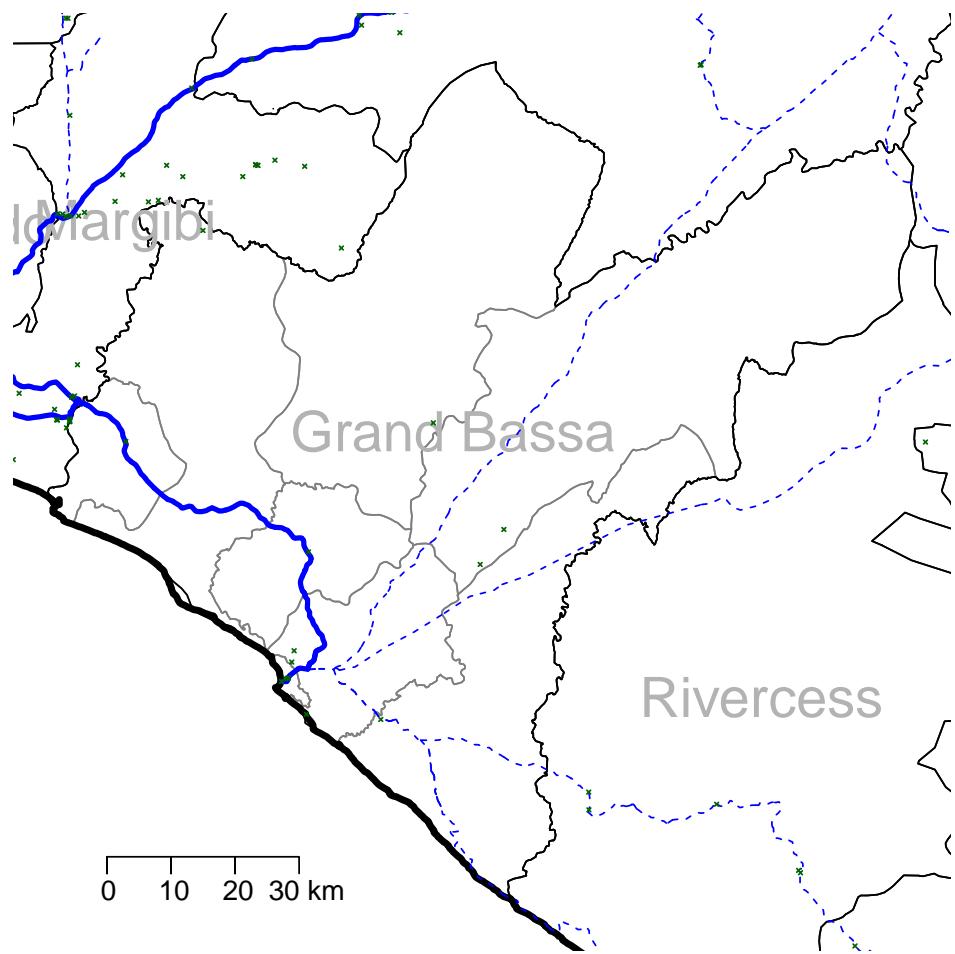


Figure 2.2: Small scale map of Montserrado and Grand Bassa in Liberia showing all districts, roads and points of interest

- 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.

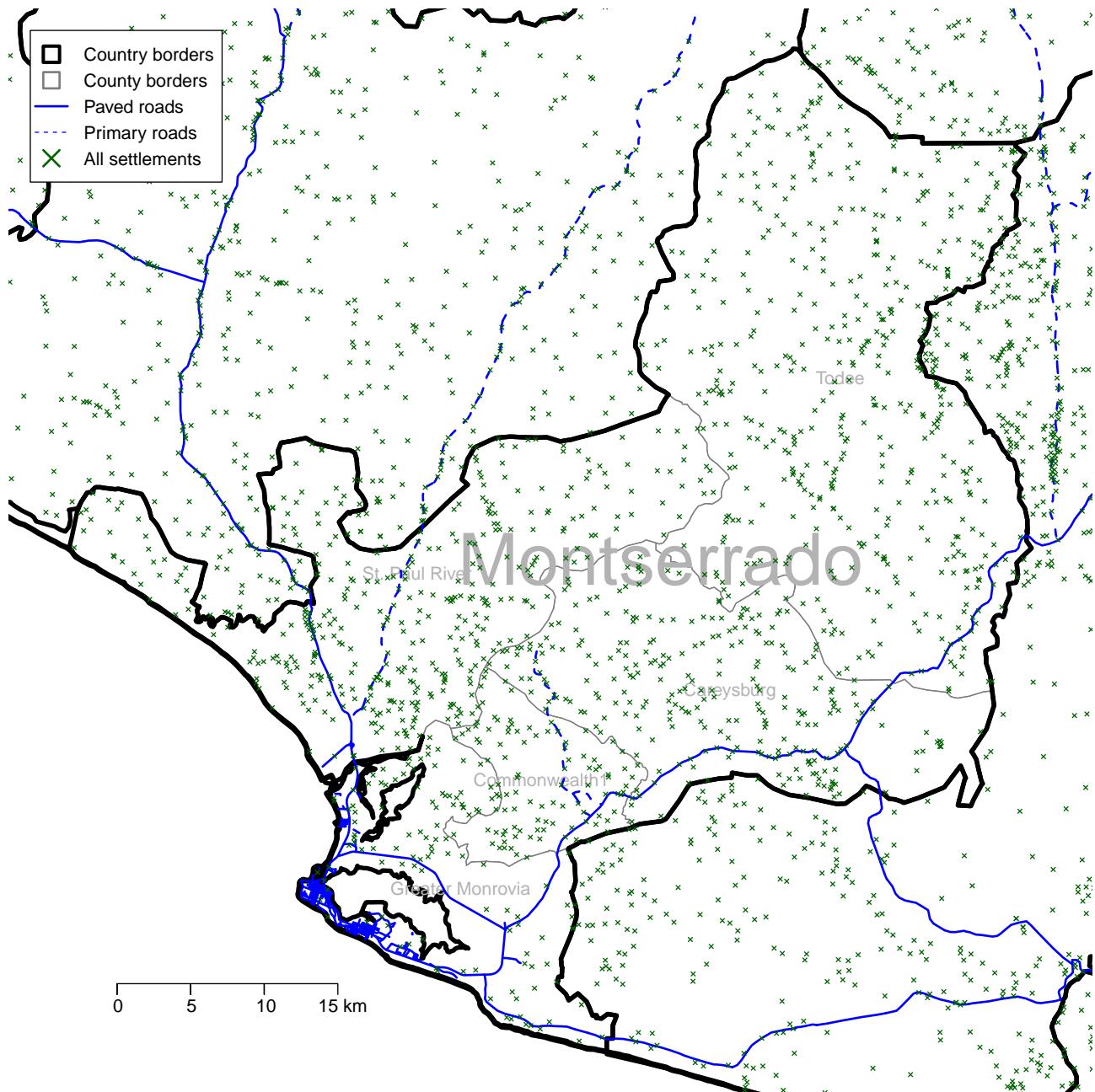


Figure 2.3: Large scale map of Montserrado county in Liberia showing all districts, roads and all settlements (towns, villages)

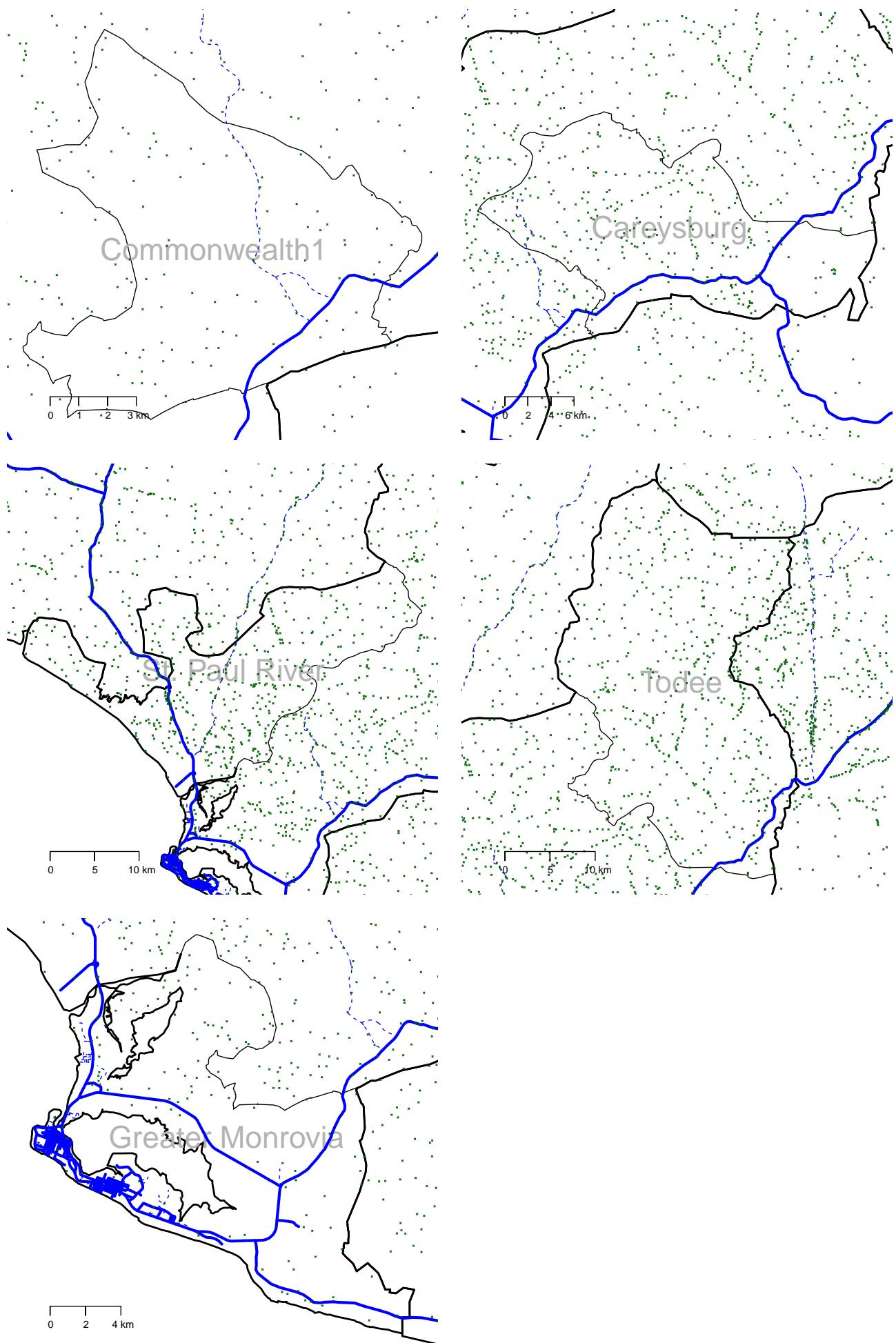


Figure 2.4: Large scale maps of 5 districts of Montserrado county in Liberia showing roads and all settlements (towns, villages)

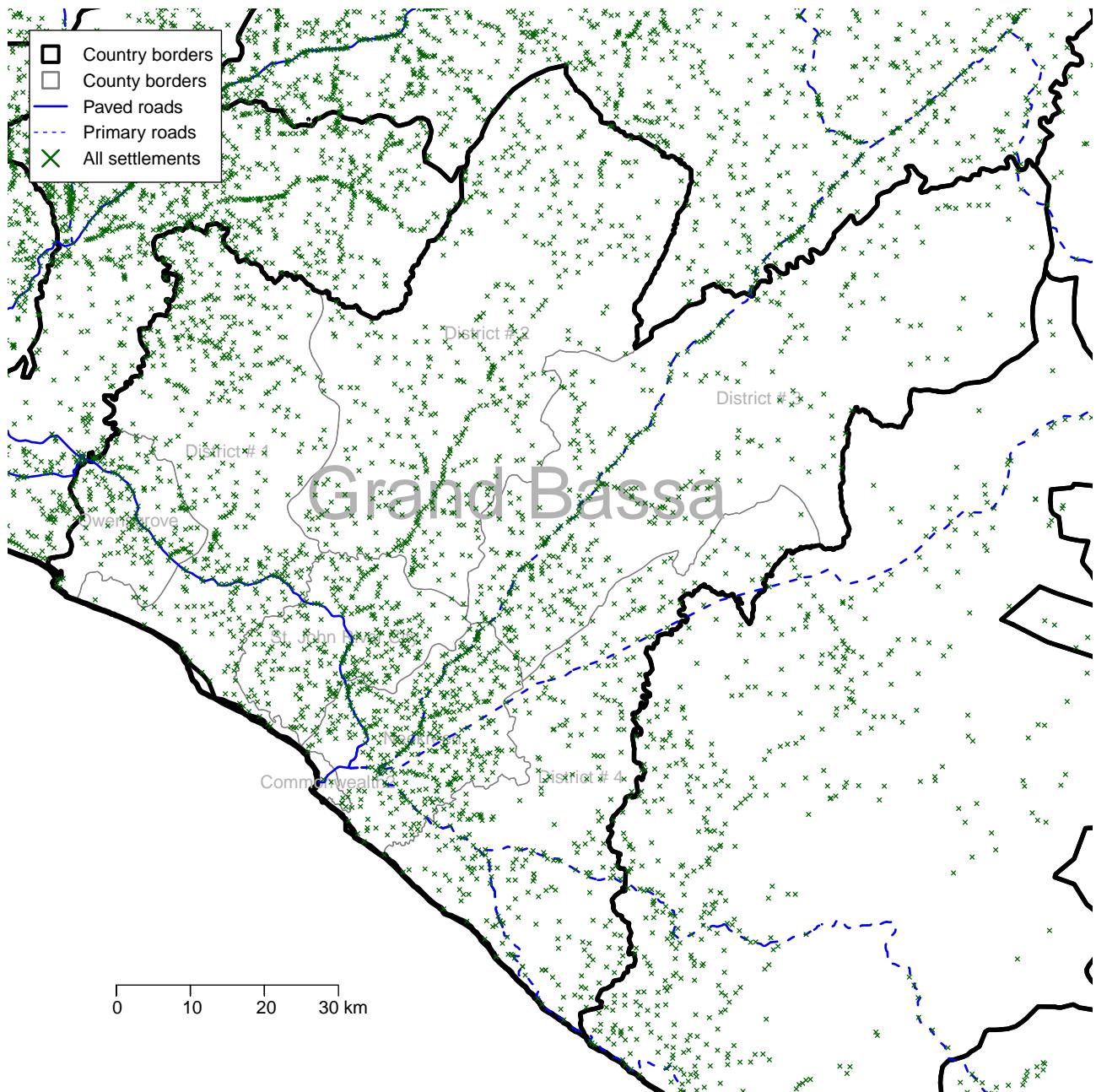


Figure 2.5: Large scale map of Grand Bassa county in Liberia showing all districts, roads and all settlements (towns, villages)

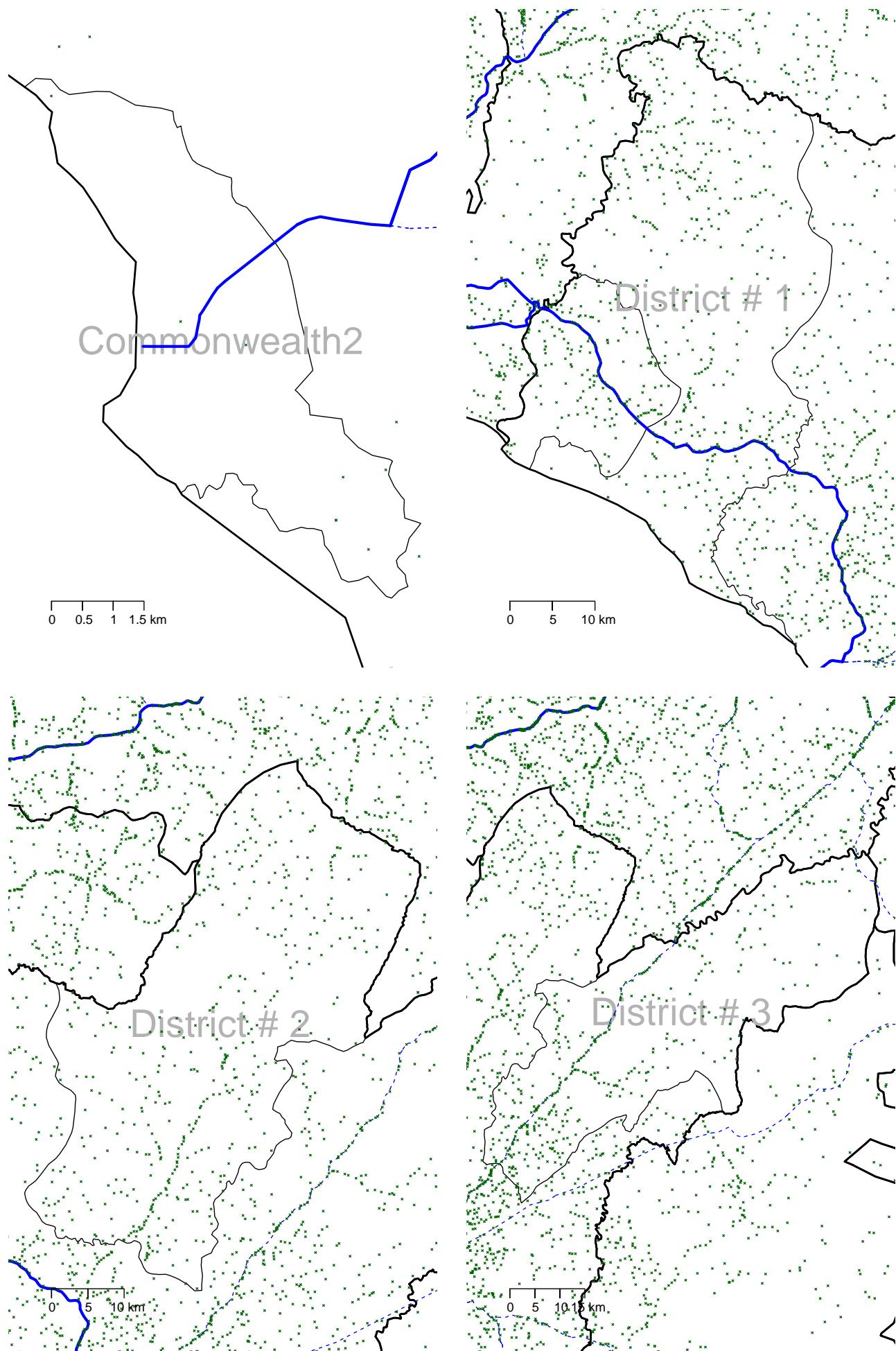


Figure 2.6: Large scale maps of 8 districts of Grand Bassa county in Liberia showing roads and all settlements (towns, villages)

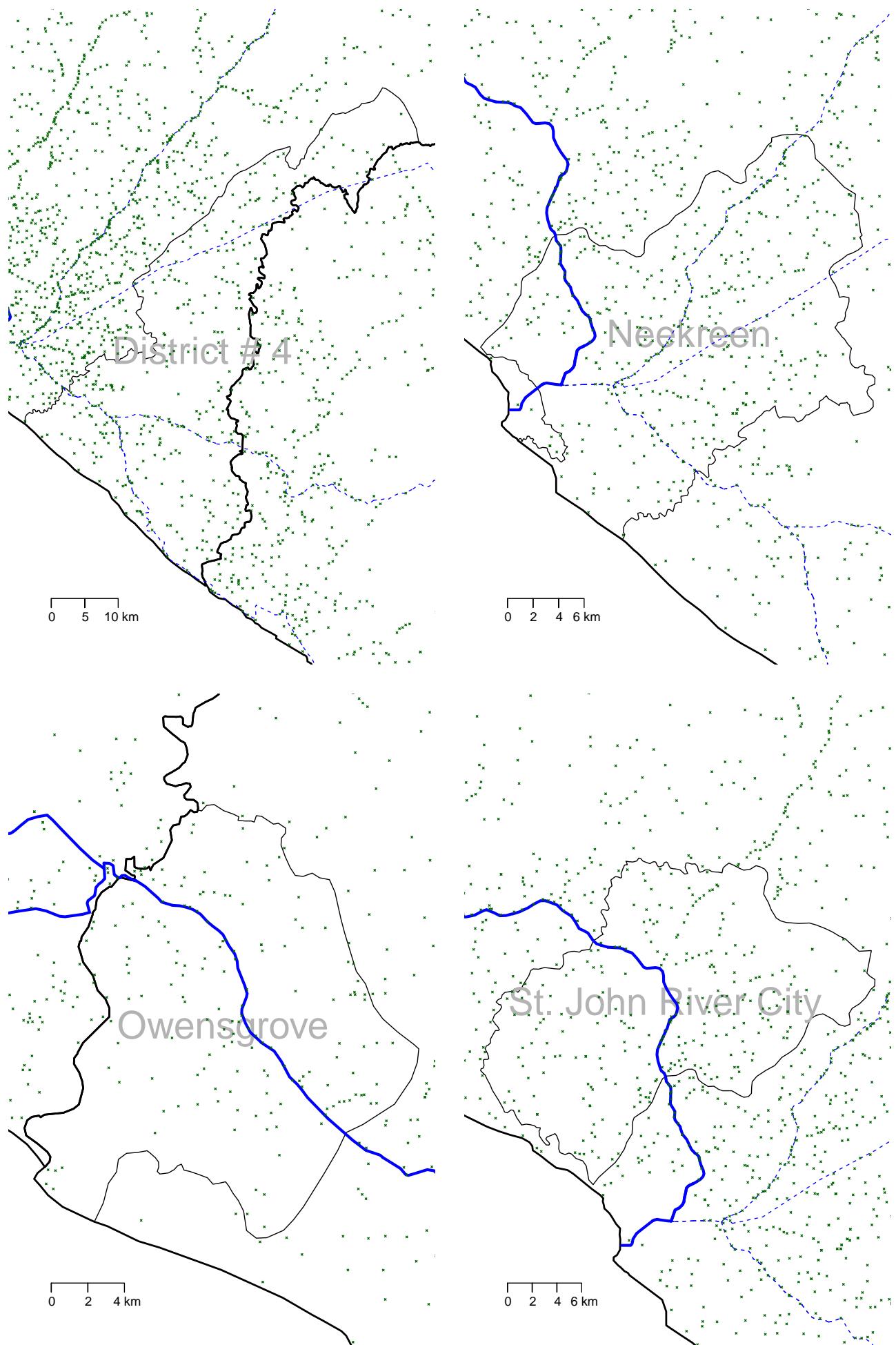


Figure 2.7: Large scale maps of 8 districts of Grand Bassa county in Liberia showing roads and all settlements (towns, villages) continued

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).

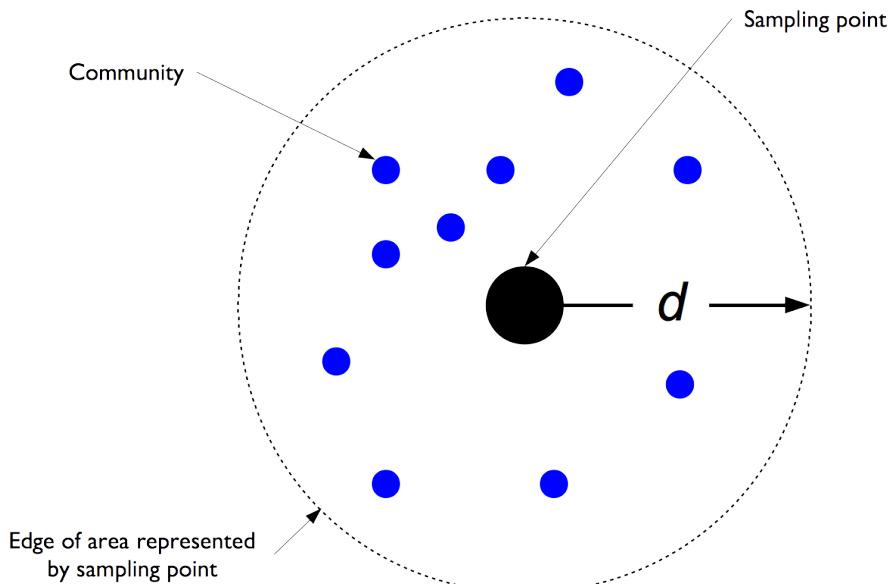


Figure 2.8: Conceptual presentation of the area represented by each sampling point

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

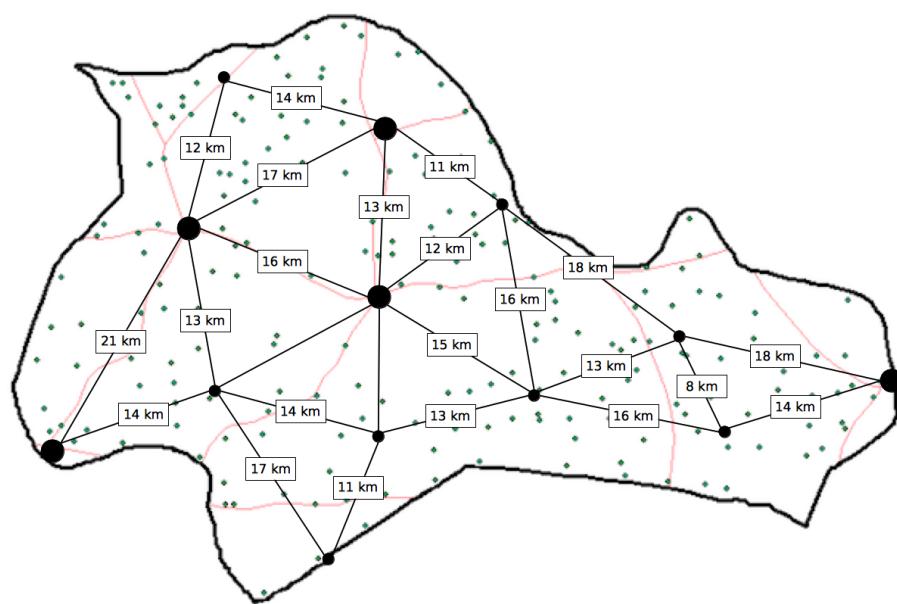


Figure 2.9: Distances of communities to the nearest substantial markets

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

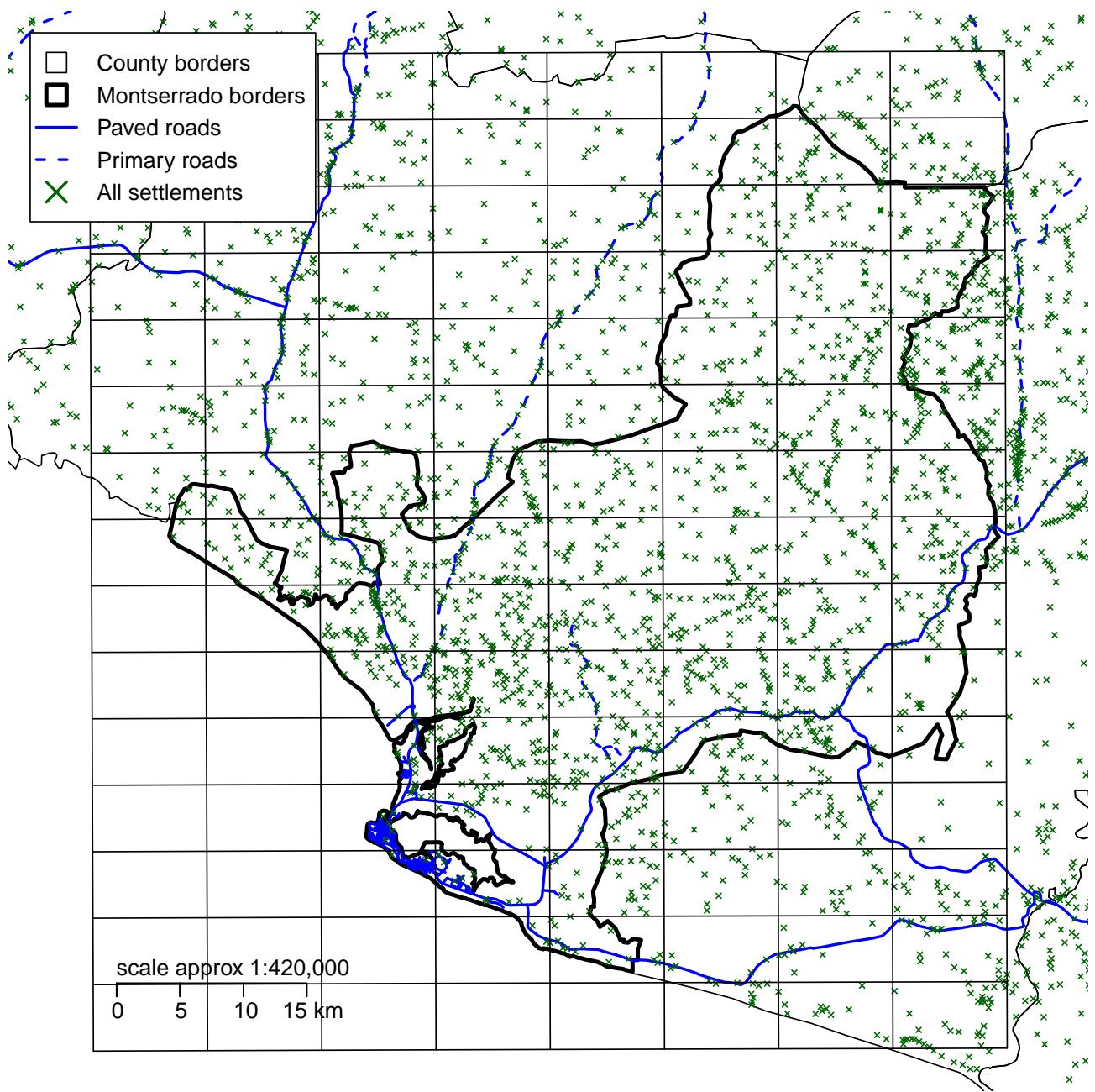


Figure 2.10: Montserrado county with a rectangular grid defined by d of 6 km

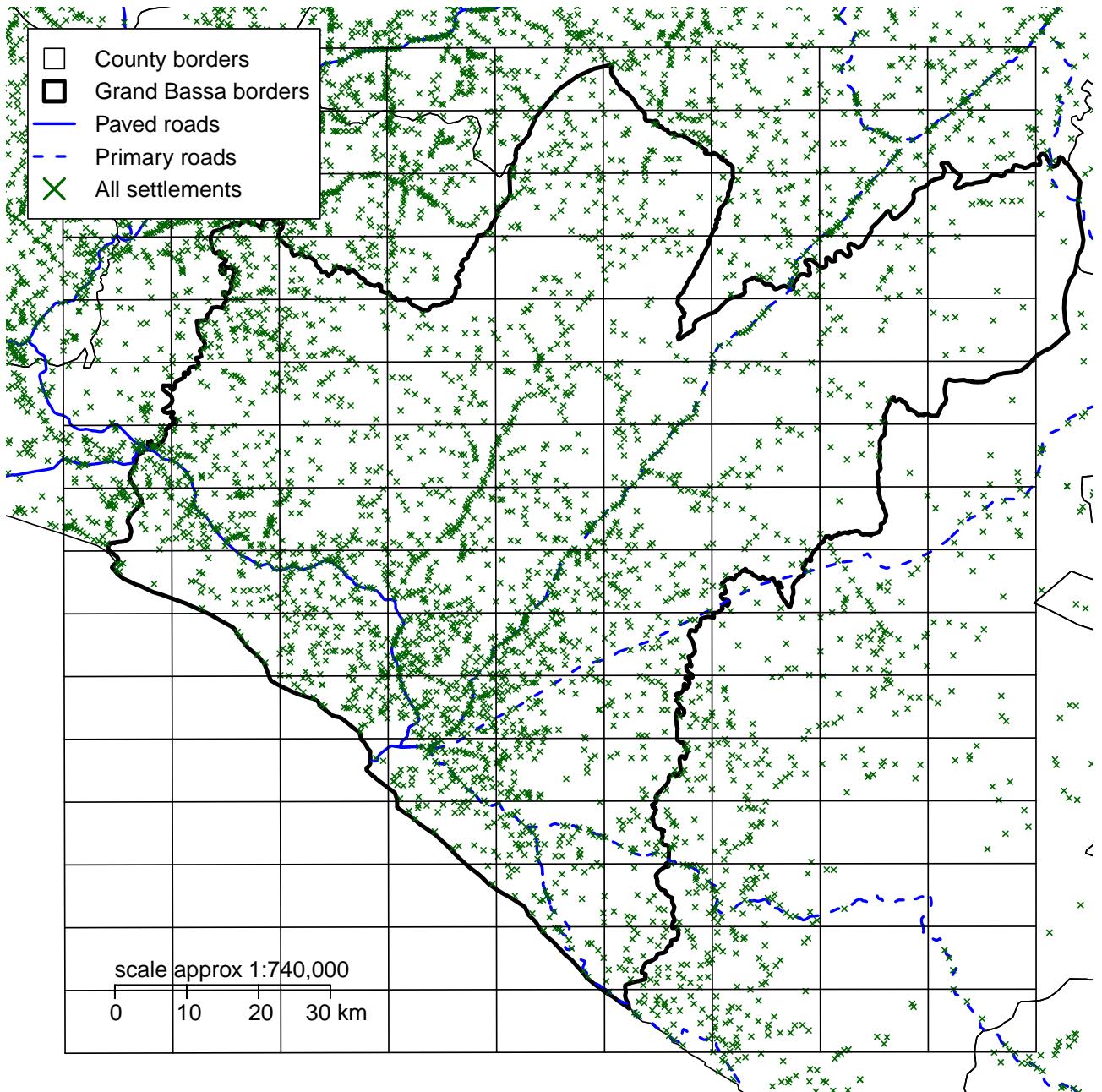


Figure 2.11: Grand Bassa county with a rectangular grid defined by d of 10 km

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

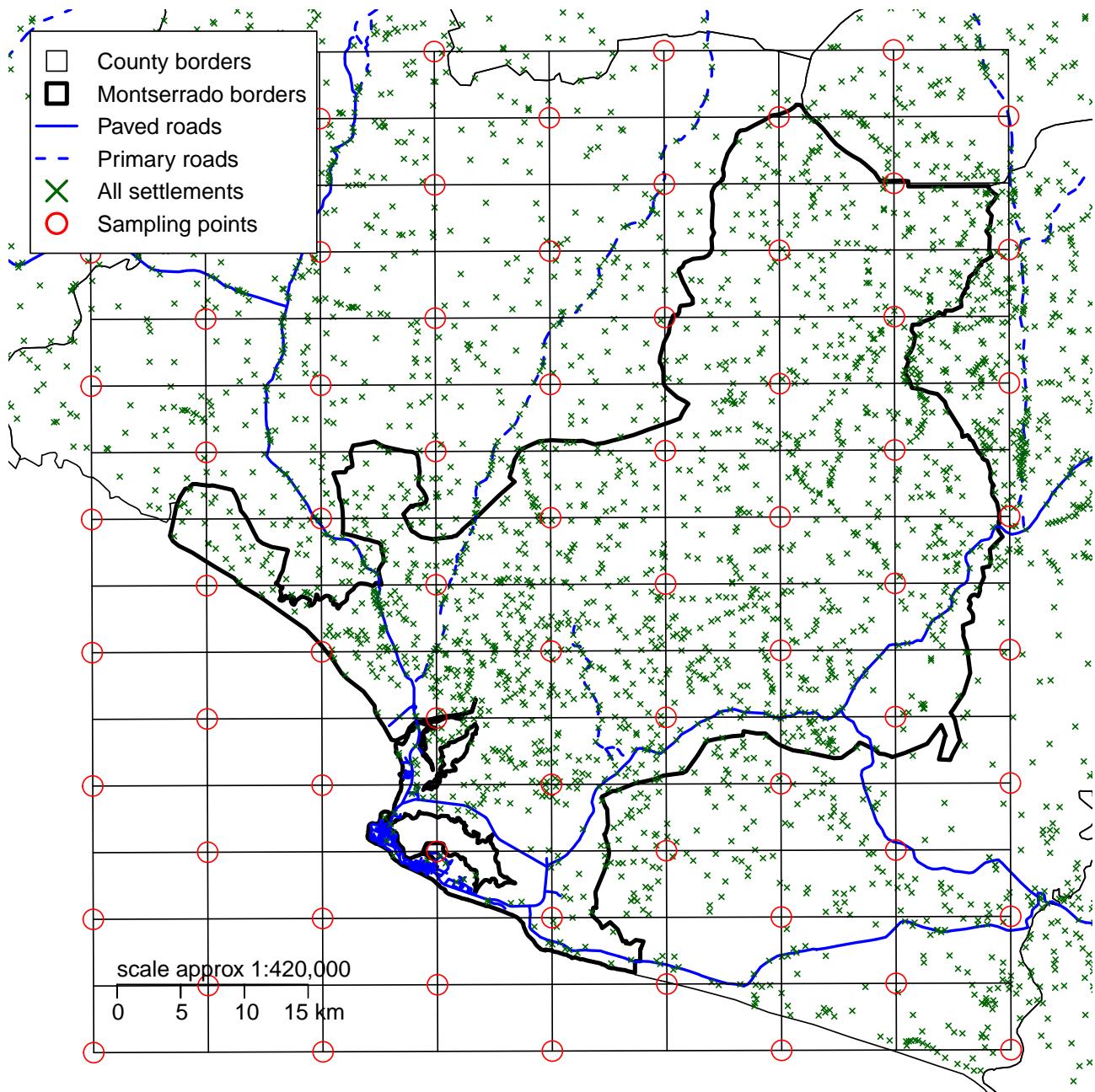


Figure 2.12: Montserrado county with a rectangular grid defined by d of 6 km and alternating intersections of the grid used to identify sampling points

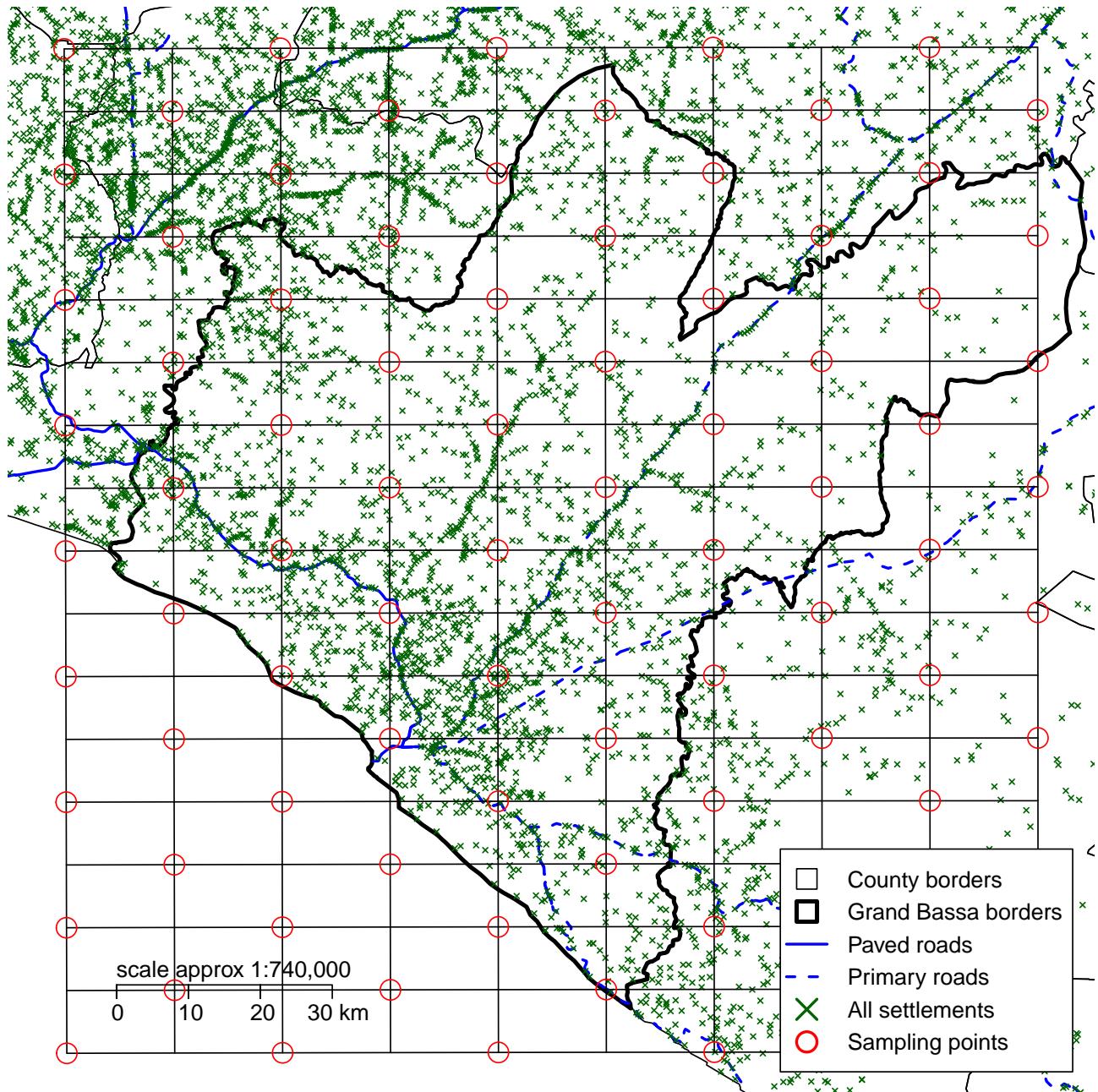


Figure 2.13: Grand Bassa county with a rectangular grid defined by d of 10 km and alternating intersections of the grid used to identify sampling points

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:

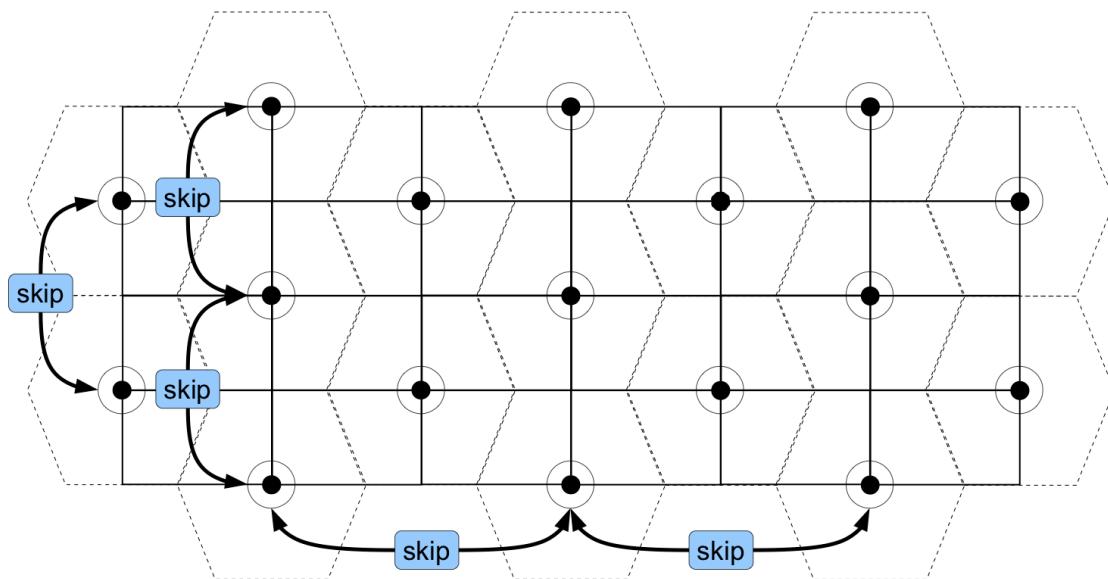


Figure 2.14: Selecting alternating intersections of the grid in the x and y directions to spread sampling points evenly

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

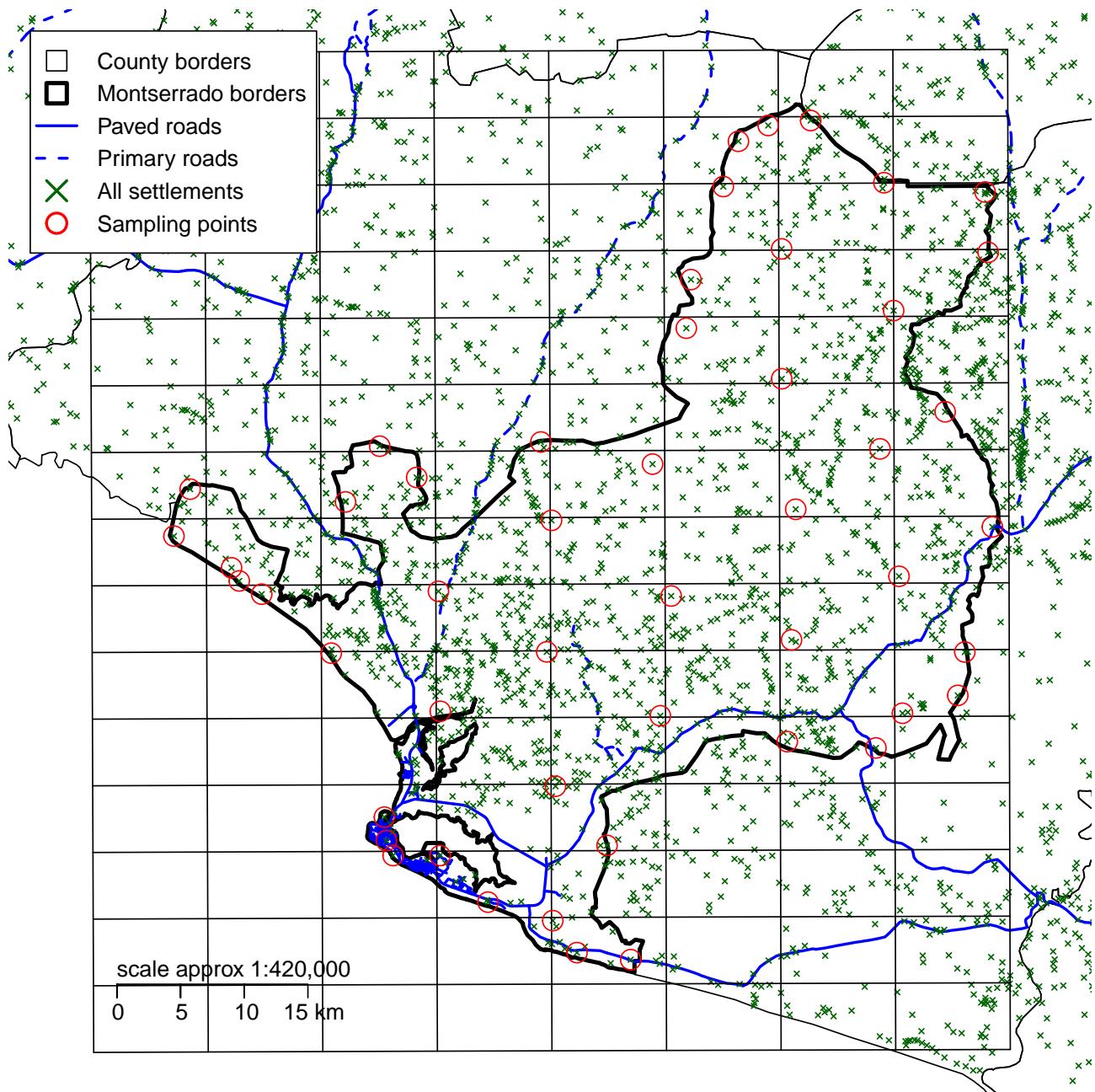


Figure 2.15: Montserrado county with a rectangular grid defined by d of 6 km and sampling points moved to the nearest communities

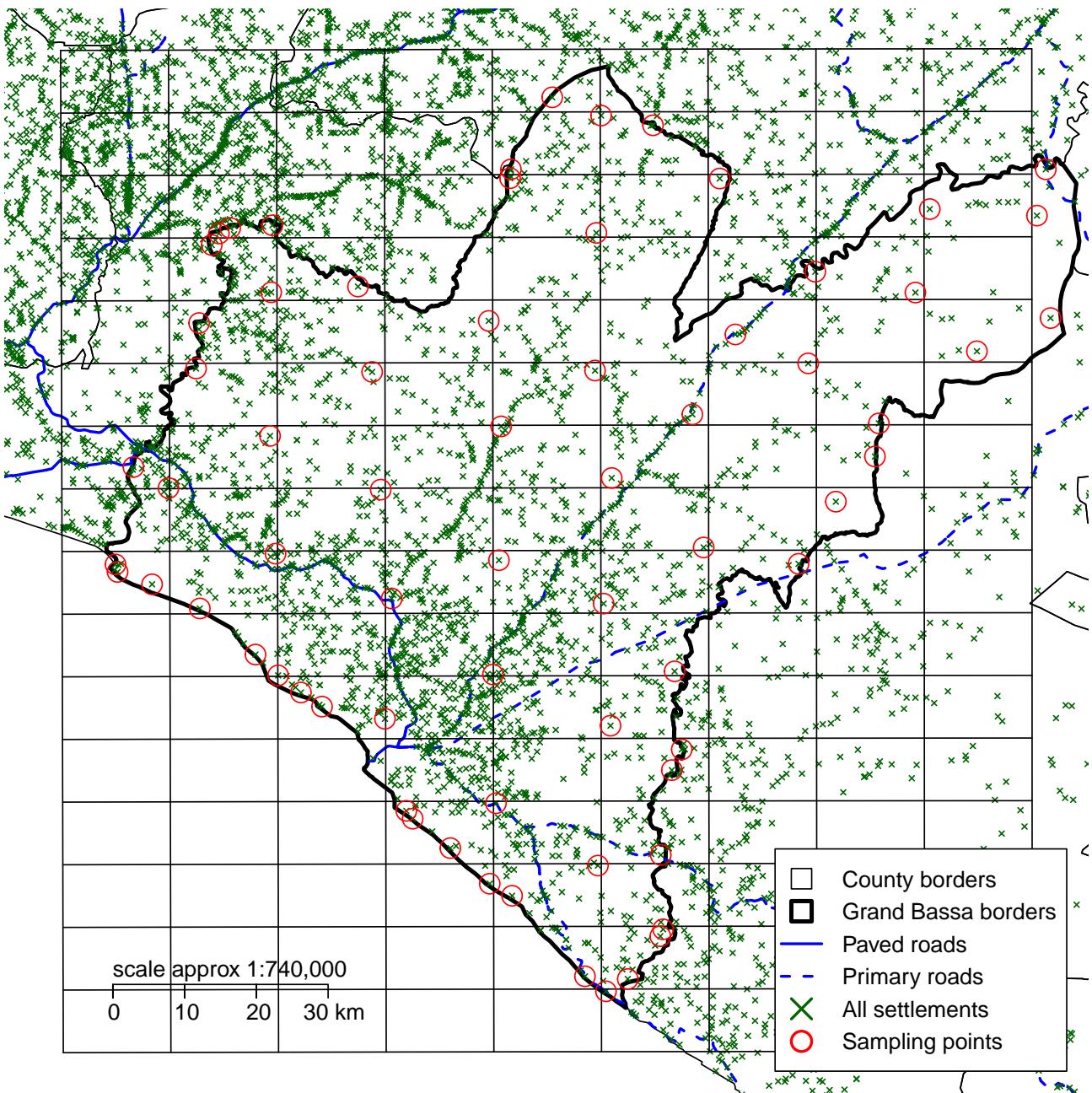


Figure 2.16: Grand Bassa county with a rectangular grid defined by d of 10 km and sampling points moved to the nearest communities

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.

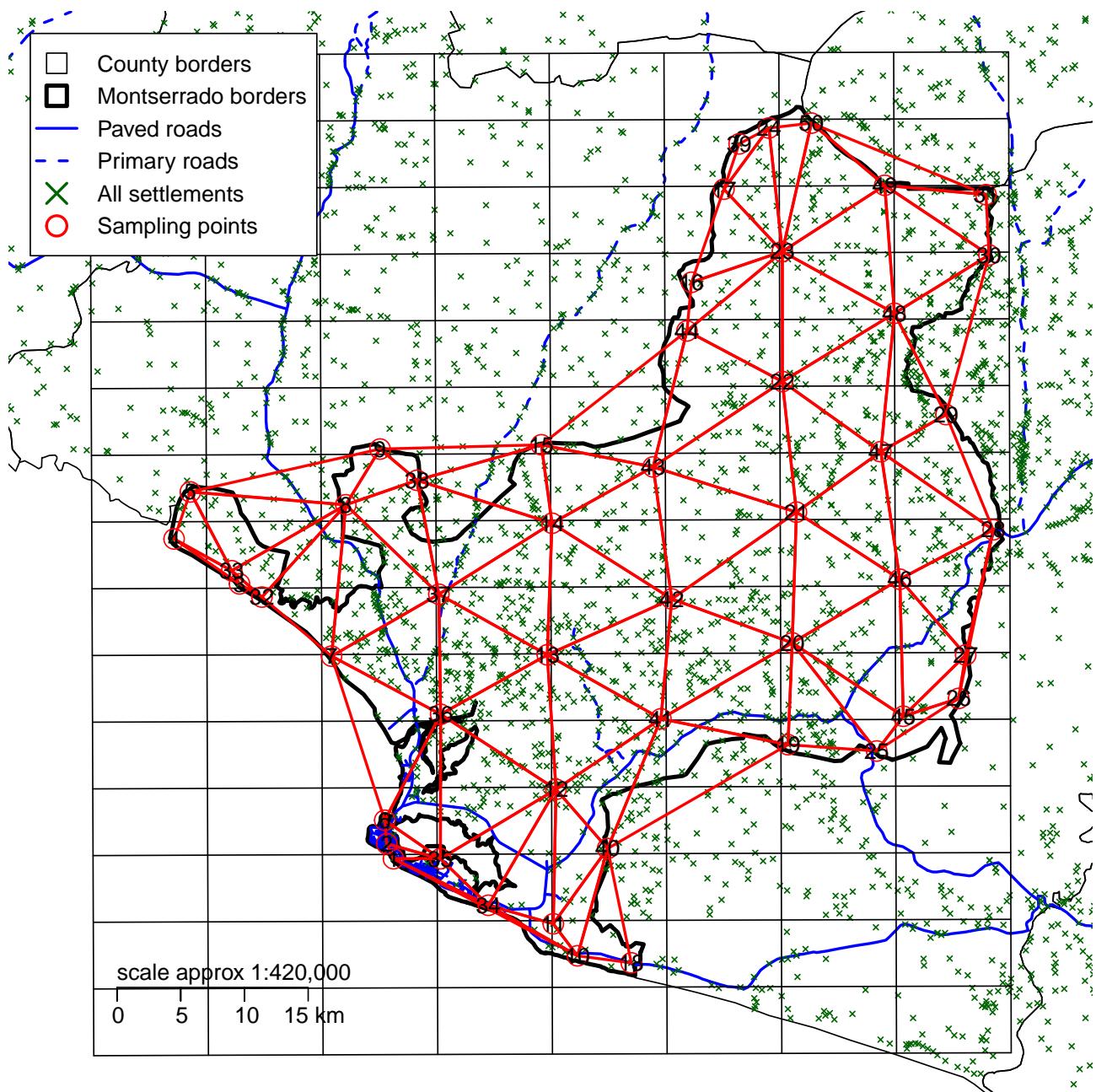


Figure 2.17: Montserrado county with a rectangular grid defined by d of 6 km and sampling points moved to the nearest communities showing test triangulation

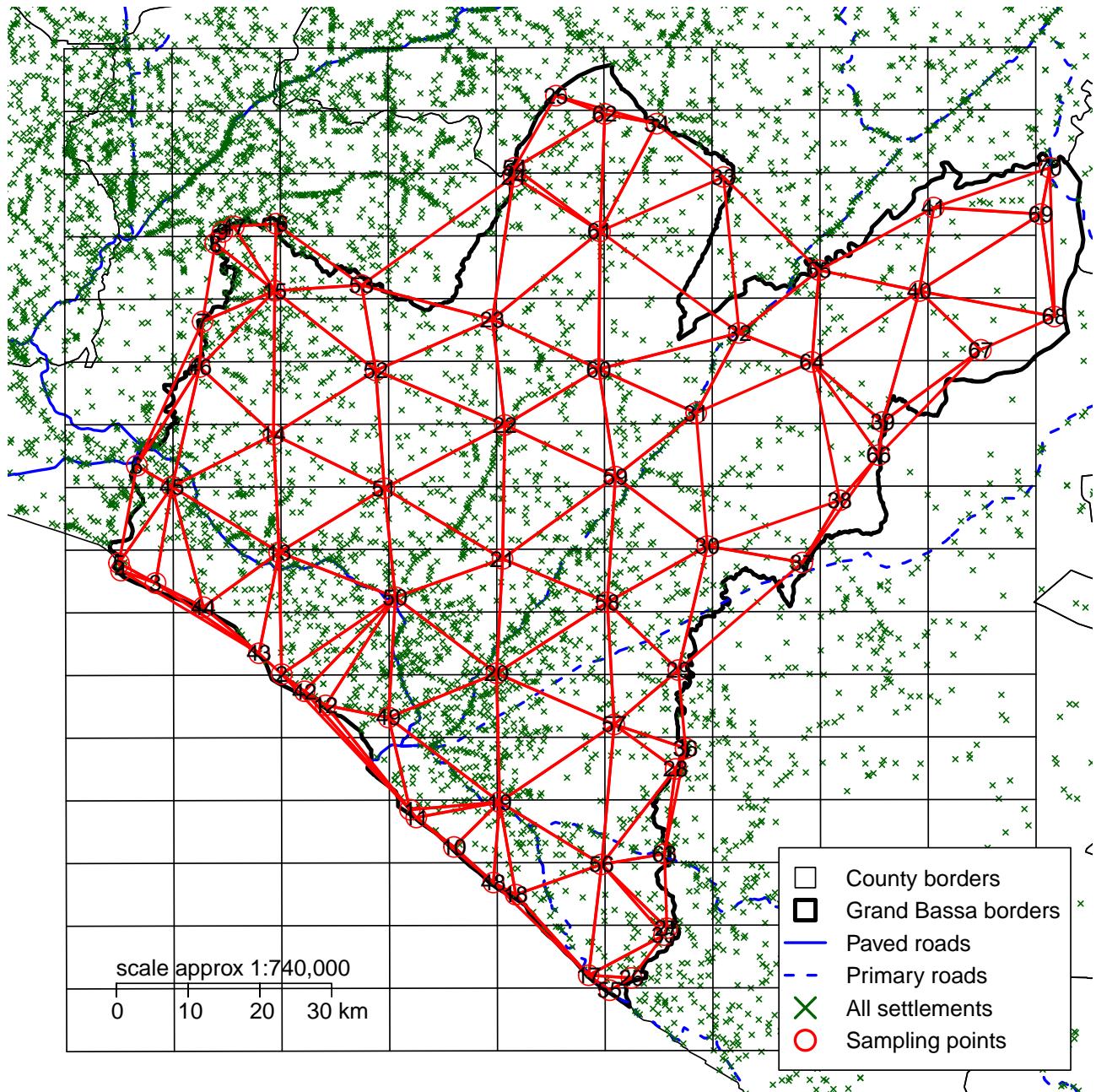


Figure 2.18: Grand Bassa county with a rectangular grid defined by d of 10 km and sampling points moved to the nearest communities showing test triangulation

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.

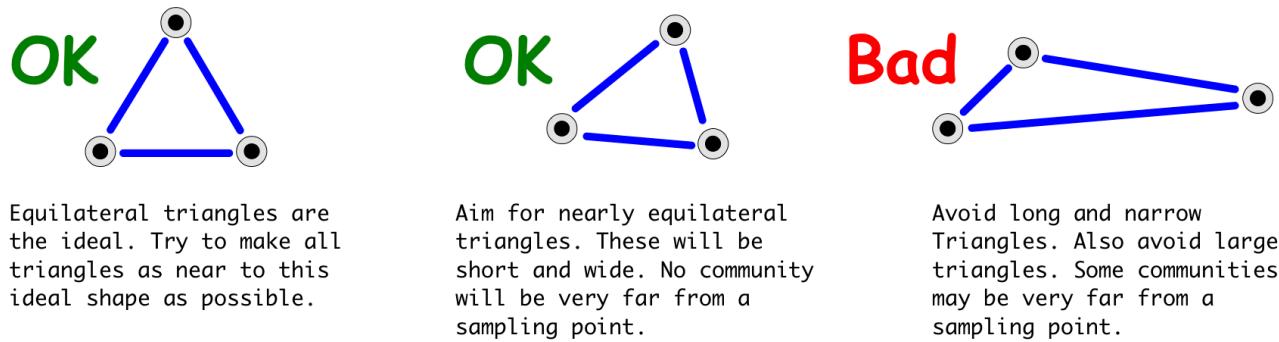


Figure 2.19: Selecting alternating intersections of the grid in the x and y directions to spread sampling points evenly

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.

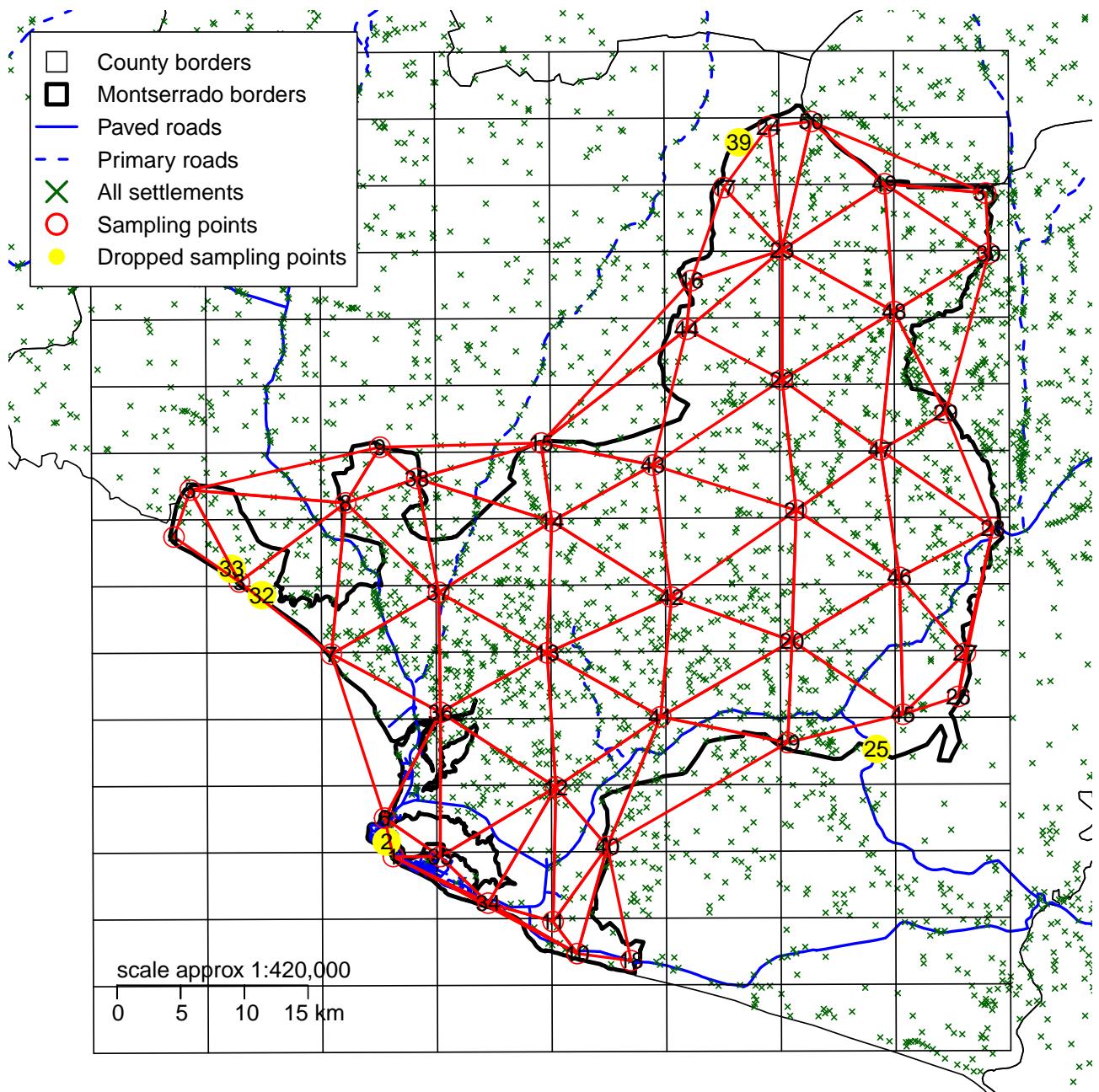


Figure 2.20: Montserrado county with a rectangular grid defined by d of 6 km and sampling points moved to the nearest communities showing updated triangulation

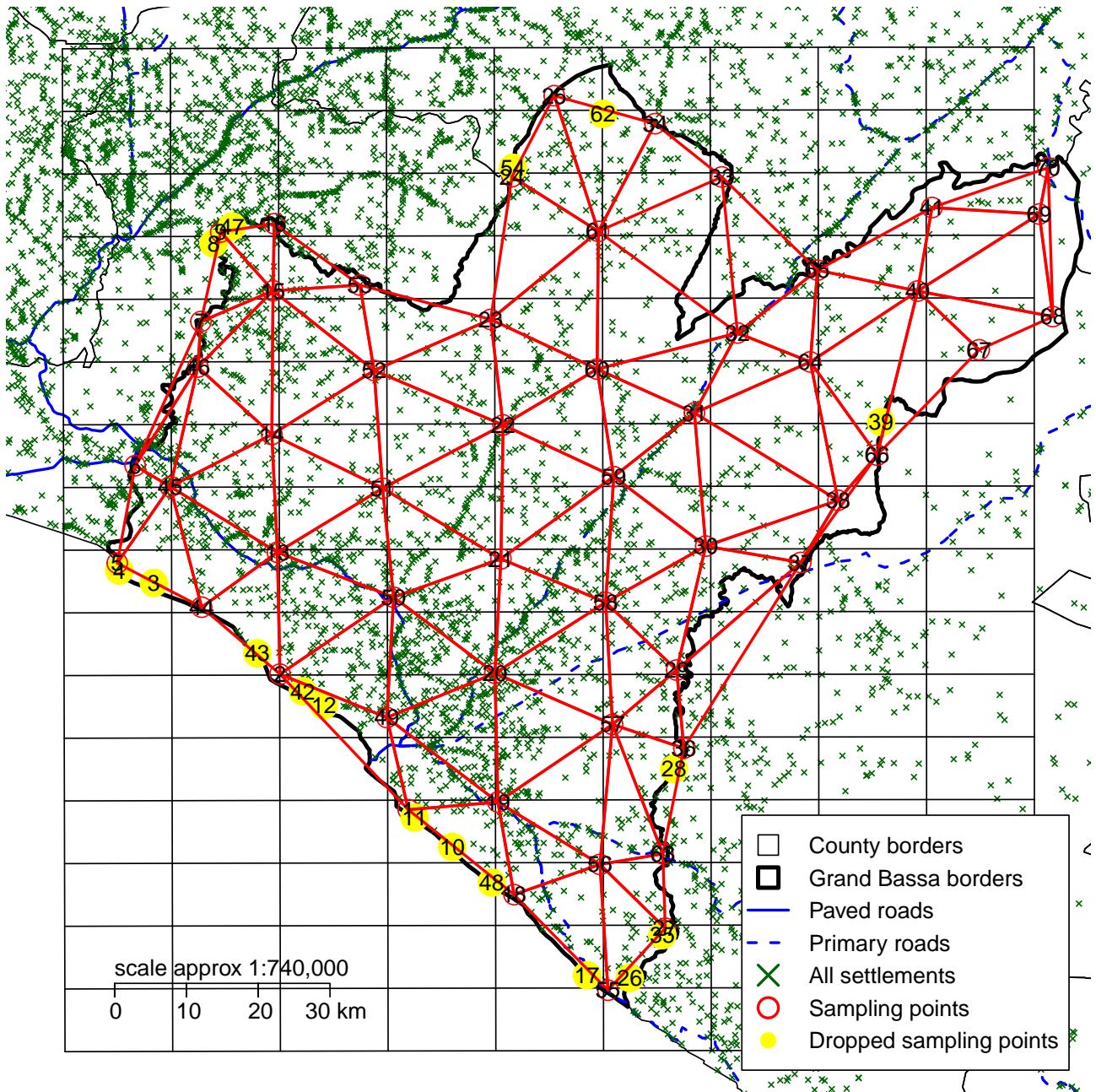


Figure 2.21: Grand Bassa county with a rectangular grid defined by d of 10 km and sampling points moved to the nearest communities showing updated triangulation

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

2.2.1 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. There will be a survey for CMAM coverage, a survey for children under 60 months, a survey for children under 24 months and a survey for pregnant and lactating women.

For the survey for CMAM coverage, an active and adaptive (snowball) case-finding method will be used in stage 2 sampling to find all or nearly all of SAM cases in the stage 1 sample.

For the under 60 months and under 24 months sample, this can be implemented in a single survey in such a way that a large enough under 60 months children are sampled to provide the minimum sample size needed for the under 24 months survey.

For the survey for pregnant and lactating women, this can be approached in a couple of ways. The standard approach to evaluating such indicators is to sample women with children less than 24 months. The expectation here is that these women are meant to be lactating given their child's age. At the same time, they have been recently pregnant (within the last 2 years) such that they will be able to recall their pregnancy experience and be able to report whether or not they benefitted from ferrous sulfate-folic acid supplementation during their latest pregnancy. The more complicated approach is to further subdivide this sample to lactating women (those with children under 24 months old) to assess the IYCF counselling coverage and then to find all or nearly all currently pregnant women to ask them about their ferrous sulfate and folic acid supplementation status. Either approaches will have their limitations. The main drawback with the second approach is that the amount of currently pregnant women at any given time is relatively small (on average about 4 in a village of about 200 population). So, finding enough sample of pregnant women will always be a challenge. This is the main reason why standard surveys such as DHS and MICS use the first approach but with women with children less than 60 months (longer recall which is problematic).

The most practical stage 2 sample for the Liberia S3M will be a survey specific for the CMAM coverage in which a part of the survey team will apply active and adaptive case finding in about 3-5 villages nearest to the chosen sampling point in stage 1. This is done so as to get enough sample of SAM cases for estimation. This CMAM coverage survey will then be nested within an overarching survey of mothers with children under 60 months to assess all other indicators (of the child and of the mother). This is the most efficient and robust approach given the indicator set required.

2.3 Sample size considerations

In general, the sample size needed for proportion-type indicators such as coverage 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.

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.

For vitamin A supplementation, micronutrient powder supplementation, IYCF counselling and ferrous sulfate-folic acid supplementation¹, the sample size of 192 would be the target sample size.

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

¹If pregnant women is chosen as the universe population for the ferrous sulfate-folic acid supplementation, the sample size requirements can be reduced given that the number of pregnant women will be finite population as well and an active and adaptive case finding method will need to be applied to find pregnant women.

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.

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:

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

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

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 program

0

Number of current cases not in the program

2

Number of recovering cases in the program

34

In this regard, 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 may have a period coverage that is misleadingly high because of high number of recovering cases when in fact point coverage is the more reflective coverage of the programme. In addition, period coverage suffers from 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.

A recent Epicentre review ([Epicentre, 2015](#)) has highlighted the existence of these two coverage estimators as a source of confusion, with the risk that implementers may choose to assess period coverage (rather than point) simply because it produces a higher estimate. 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. For this reason work has been conducted on further developing the coverage estimators to address this confusion around them. This work has included:

- A shift in terminology 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 MAM and then getting them to treatment. A programme with effective case-finding will always have good overall coverage when this estimator is used. **Period coverage** is now named *treatment coverage* as this is the estimator that approximates this coverage indicator the closest.
- Improvement of the period coverage estimator so that it can be used more precisely as a single coverage estimator for treatment coverage.

We propose therefore that programmes report both case-finding effectiveness and the improved treatment coverage indicator and, crucially, give sufficient context to properly evaluate both estimates. For a more detailed description of the definitions of these two estimators, their differences and limitations, and the development work completed under this project, see Annex 3.

Chapter 4

Questionnaire

Chapter 5

Analysis

Bibliography

Epicentre (2015). Open review of coverage methodologies: Questions, comments and ways forward. Technical report, Epicentre.

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.