

# **Sampling Plan for AMTSL Surveys**

## ***Introduction***

The primary objective of the AMTSL surveys is to calculate the proportion of deliveries at health facilities at a national level that receive AMTSL. Another objective is to measure characteristics of health facilities where AMTSL is provided, including the content of treatment guidelines, in-service training, and availability of relevant drugs.

A source of data is required to estimate these indicators. What data should be used to calculate these indicators? With sufficient time and resources, a census of deliveries could be done to get the actual values for these indicators. Such an effort is not practical given time and budget constraints. Reasonable estimates of these indicators can be found by sampling the populations of interest. Conducting a survey with a sample of the population, rather than a census, greatly reduces the time and cost of data collection and analysis required for these indicators.

How should we implement this sampling process? The first step in the sampling process is to determine the size of the sample required to produce reasonable estimates of the primary indicators of interest. In our surveys the primary indicator is the proportion of deliveries that receive AMTSL. This is the first step because the sample size directly affects the cost of the survey. Then we must know the population from which the sample will be selected. When that population is known we can proceed to select the sample. After selecting the sample, weights are calculated to correct for errors in the sample. The sampling process is not perfect, both by design and due to errors that occur in selecting the sample. Weights are incorporated into the analysis to correct for these errors in the sample.

The sampling plan that follows gives more details on how to calculate the sample size, determine the population of interest, select the sample from that population, and calculate weights to correct for errors in the sample.

## ***Requirements for Implementing the Sampling Plan***

Before proceeding with implementing the sampling plan, consideration should be given to who will be responsible for implementation. The person in charge of sampling should be someone with prior sampling experience, preferably with surveys of health facilities or surveys of households and individuals within those households. Ideally that person will have a statistics or biostatistics background, or some other field that does training in surveys and sampling. While the design of the plan is meant to be simple and straightforward, having someone with statistics training and sampling experience will greatly aid in the process.

In addition to personnel requirements, sufficient background data are required to draw the sample. These data should include a list of all hospitals that provide delivery services, the location of those hospitals, and an estimate of the number of deliveries in the preceding year. Hospitals may record number of births, not deliveries. In that case it is

preferable that the number of births be the number of live births, plus still births, which should be equivalent to the number of deliveries. If only live births is available that will serve as a proxy for number of deliveries. The estimate of number of deliveries should come from as complete hospital record data as possible, because this estimate will be critical for determining if there are enough deliveries per day to make it worthwhile to include the hospital in the sample. The average number of deliveries per month can substitute for number of deliveries per year (ie. the average number of deliveries per month is the number of deliveries per year divided by the number of months (12)). If neither the number of deliveries per year nor the number per month are available, the number of deliveries in the last month can serve as a proxy for these. This may require asking someone to go to the labor and delivery room register to obtain a count of the number of deliveries in the last month if this information is not available via routine sources.

In the above paragraph and elsewhere in the text we refer to deliveries, as if all deliveries were included in the sample, but that is not the case. Only deliveries where AMTSL could occur are included in the sample, because the objective is to observe whether AMTSL occurred for these deliveries. AMTSL can be performed for any non-surgical vaginal deliveries, which includes those performed by vacuum extraction and forceps. Cesarean sections will not be in the sample. If information were readily available on cesarean sections, the number of deliveries mentioned above would exclude the number of cesarean sections. Where this information is not available the number of deliveries or births will provide information for sample selection. Not accounting for cesarean sections would only be problematic in countries with a high number and variability in cesarean sections across facilities. Cesarean sections would not affect the sample in countries with a relatively constant rate of cesarean sections across facilities.

Throughout this sampling plan we will refer to hospitals as the health facilities included in the sample. In some countries, health centers might be included as well. This would be for countries where health centers are allowed to provide AMTSL and some of these health centers have sufficient deliveries to be included in the survey. Please note, ideally, it is preferable to include hospitals and health centers in the sample, if possible. This will allow one to investigate use of AMTSL at lower volume health facilities. However, where the number of deliveries is small, it has generally not been found to be practical to include health centers in the sample. Your observers will spend too much time waiting and may observe no deliveries at all during the prescribed observation period.

### ***Sample size***

The goal of determining sample size is to find a size that fits within the desired range for precision and within the budget available. Often the initial goals for precision result in a sample that is beyond the funds available for the survey. One way to address this problem is to reduce the level of precision to make the survey within budget. Such a solution is reasonable as long as the level of precision chosen is meaningful relative to the indicators to be estimated.

Level of precision is not the only factor that affects the size of the sample. The way time enters into the calculations also has an effect. Survey samples can be designed to

measure indicators at one point in time (point estimates) or to measure change in the indicator over time. The point estimate approach produces smaller sample sizes than the change over time approach. Those designing surveys have to take into account whether measuring change over time is worth the additional cost of a larger sample, or whether a point estimate will be sufficient.

The initial calculations for the PATH-sponsored AMTSL surveys examined both the point estimate and change over time approaches. To date, it has not been possible to measure change over time given the funds available for these surveys. The sample size required for point estimates could fit within the budget. Countries with more funding for the surveys could opt to design the surveys to measure change over time. In the initial set of surveys that measured AMTSL prevalence the target sample size was 200 deliveries. This sample size was determined based on a point estimate with a precision of 10% on either side. In order to make this more robust, the target sample would be 213, which is the largest sample size for a point estimate with 10% precision (see Table 1, 10% precision, 50% prevalence).

The approach and sample size selected for a survey could vary from the recommendation, although it is unlikely that a smaller sample would be appropriate. Tables 1 and 2 are provided to facilitate determination of sample size based on assumptions different from those given for the sample size recommended above. Table 1 is for point estimates and Table 2 is for change over time. If prevalence and precision/change over time are different from than those found in Tables 1 and 2, calculate sample size with the equations found in Appendix A.

**Table 1. Sample size for a given prevalence and level of precision at a single point in time**

Precision <sup>1</sup> (+/- %)	Prevalence										
	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%
3%	854	1,209	1,518	1,779	1,992	2,158	2,277	2,348	2,371	2,348	2,277
5%	307	435	546	640	717	777	820	845	854	845	820
10%	77	109	137	160	179	194	205	211	213	211	205

<sup>1</sup>Precision is the distance from the estimated prevalence in either direction.

To select a sample size from Table 1, look across the top for the value of AMTSL prevalence that is close to what is expected in your country, then move down that column to a row with the desired precision. In this case precision is the amount of error above and below the estimate that is considered acceptable. For example, if the proportion of deliveries receiving AMTSL is 35, and the desired precision is 5%, the sample size would be 777. The desired precision is 35% plus or minus 5%, which is a range from 30% to 40%.

**Table 2. Sample size for a given prevalence and amount of change over time**

Change <sup>1</sup>	Prevalence										
	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%
-10%	169	286	389	478	554	615	663	697	718	725	718
-5%	843	1,281	1,664	1,992	2,266	2,484	2,648	2,758	2,813	2,813	2,758

<b>-3%</b>	2,506	3,691	4,724	5,605	6,334	6,911	7,336	7,610	7,731	7,701	7,519
<b>3%</b>	3,235	4,329	5,271	6,061	6,699	7,185	7,519	7,701	7,731	7,610	7,336
<b>5%</b>	1,281	1,664	1,992	2,266	2,484	2,648	2,758	2,813	2,813	2,758	2,648
<b>10%</b>	389	478	554	615	663	697	718	725	718	697	663

<sup>1</sup>Change is in percentage points relative to prevalence.

Selecting a sample size from Table 2 follows a similar process. Look across the top for the expected baseline prevalence, then down the column for the row with the desired change in prevalence. Note that increases and decreases in prevalence do not give the same values. An example is if the estimated baseline prevalence is 35% and the prevalence is expected to increase by 5 percentage points, the sample size would be 2,648. If it decreases by 5 points the sample size would be 2,484.

Based on these two examples it can be seen that the point estimate has a much lower sample size than change over time (777 verses 2,648). Other examples give the same result. For example, start with 20% prevalence and design two surveys (baseline and follow-up) to measure a five percentage point drop in prevalence. Based on point estimates, the baseline survey would have a sample of 546. See the 20% column and row for 5% precision. The follow-up survey would have a sample of 435, which can be found in the 15% column and 5% precision row. In this case, the change over time estimate in Table 2 is 1,664. The sample size in this case can be found in the 20% column and -5% row. The point estimate is again lower than the change over time estimate (435 verses 1,664).

### ***Sample Population***

In order to select a sample, we must know the population from which the sample will be selected. The population is important because that is the group of items that the indicators must represent. The item in our case could be a person or health facility and the group would be all those items that fit within certain criteria. When the population is known, we can devise methods to select a sample that will properly represent the whole population. Conversely, if the sample is poorly selected, indicators calculated from that sample would not represent the population well due to biases in the sample. Even well constructed samples can have known biases, some of which can be corrected with weights. Ideally errors in the sampling that are not corrected will be small enough not to make a difference in the indicators calculated from the sample.

Normally the population is all those items that are members of a group at a given point in time. For example, in a survey of individuals, the population would be those individuals of the desired gender (women and/or men), age range (women and/or men of reproductive age) and geographic area of interest (country or selected areas within a country). The item is the individual and group membership is determined by whether the individual is of the proper gender, age and country. In a survey of health facilities, the population would include those health facilities that provide the services of interest. The item in this case is the health facility and group membership is determined by whether the facility provides the service of interest. In our case, a health facility survey will include

those health facilities for which there is some evidence or expectation that AMTSL is provided.

The population for the indicator of AMTSL prevalence is different because the population of interest occurs over time and not at a single point in time. At any instant in time there may or may not be deliveries occurring at a given facility. These deliveries accumulate over time. What then is the right time period for the population who might receive AMSTL? Could it be any arbitrary time period? The answer to these questions relates to the nature of how deliveries occur over time. The number of deliveries have been shown to follow seasonal patterns. More deliveries occur in some seasons than others. One factor that can affect seasonality of deliveries is seasonal migration related to work. Due to the seasonality of births, a one year time frame is appropriate for the population of deliveries, because one year covers all the seasons.

Surveys, however, do not occur over such a long time frame. Surveys occur over a relatively short time frame. The implication is that even if the sample of facilities provides a representative sample to estimate AMTSL prevalence in the short term, the seasonality of deliveries could make the sample less representative over time. While this problem may not be easy to correct, it should be kept in mind when interpreting data, and when planning follow-up surveys, in countries that decide to conduct baseline/follow-up surveys.

### ***Selecting a sample***

Various approaches are taken when selecting a sample. The most complete is a census, which is not a sample at all, because it contains all the individuals/health facilities of interest. The advantage of a census is that it is fully representative of the population from which it was selected. The disadvantage is cost. A lower cost alternative is a simple random sample, where individuals/health facilities are randomly selected throughout the country. While this is not as costly a census, a simple random sample can be expensive due to the travel time associated with the random dispersion of individuals/facilities across the country.

An alternative to a simple random sample (SRS) is to create a sample that selects multiple individuals/facilities from small geographic areas called clusters or enumeration areas (Lemeshow et al. 1990, Hulley and Cummings 1988). In a typical individual-level survey, enumeration areas (EAs) are first selected, then households are selected among those EAs and finally individuals are selected within the households. Rather than being spread across the country, households and individuals are clustered together in the EAs. The geographic proximity of the individuals reduces the cost of the survey. On the other hand, clustering produces a sample that underestimates the population variance relative to a simple random sample, which means a larger sample must be taken to offset this. The degree that a cluster sample underestimates the SRS variance is known as the design effect. Typical values for the design effect are about 2.0. The sample size must be multiplied by a factor of 2.0 to account for the design effect. The overall effect of clustering should be a lower cost survey. Even though the design effect increases the sample size, the reduction in travel time due to clustering should more than offset this increase.

The sampling approaches we propose are hybrids of these types. Because our population for deliveries spans a one year period, we could not possibly do a census of deliveries. We, however, could include all health facilities with deliveries. Even this is not quite feasible, because some facilities have too few deliveries. The compromise is to exclude those facilities with less than a given number of births: generally those with less than one delivery on average per shift. This approach is relatively simple because no random selection of facilities is required. In countries that have many health facilities with one or more deliveries per shift, it may be necessary to select a sample of facilities. Sampling facilities makes this second approach more complicated to implement and adjust for errors than the first.

### **Duration of Observation in Facilities**

Before describing the two proposed approaches, some comments should be made on the duration of observation of deliveries in facilities. The duration is the length of time spent in facilities. In household/individual surveys, there are normally a fixed number of households selected in a cluster. That works fine because people can be contacted in a reasonable amount of time and estimating the overall duration to spend in the cluster is relatively straightforward. If a fixed number of deliveries were to be observed in each facility, estimating the length of time that interviewers would spend in each facility would be difficult. We can estimate the length of time based on the average number of deliveries per month, but deliveries do not occur uniformly over time, which would make our calculations inaccurate. A more reasonable method is to fix the amount of time spent in each facility. When that is done, calculating the overall time spent in facilities and the related budget can be done with a fair amount of accuracy. It also helps with the logistics of the survey.

### **Selection by Number of Deliveries per Shift**

The first approach for sample selection involves including all hospitals with more than one delivery per shift. The way we describe such a sample is that the population of interest is all hospitals with more than one delivery per shift. Occasionally a different threshold of number of deliveries (higher or lower) is selected to ensure the survey includes observations of the required number of deliveries and no more. The approach to finding out which facilities have more than one delivery per shift is relatively straightforward. As mentioned earlier, this process requires a list of health facilities with name of facility, location of facility, and number of deliveries per year or month at each facility. The number of live births may be substituted for number of deliveries as mentioned above. The steps in the process appear in Box 1. These steps are designed to be completed with the aid of a spreadsheet.

As mentioned earlier, only non-surgical deliveries will be observed. In countries where the rate and variability of cesarean sections is high, excluding cesarean sections from the number of deliveries is important. Where this information is not available in routine sources, manual review of delivery room registers may be required to obtain this information.

<b>Box 1. Steps to select a sample based on number of deliveries per shift</b>
--

1. Calculate the number deliveries per day in each health facility
2. Calculate number of deliveries per work shift
3. Sort facilities by number of deliveries per work shift
4. Calculate the cumulative number of deliveries per shift
5. Calculate the cumulative number of deliveries per interviewer day
6. Calculate the cumulative number of deliveries per observation period
7. Make a preliminary determination of the facilities that will be in the sample
8. Adjust the calculations and the sample as needed

The first step is to calculate the number of deliveries per day. To calculate the number of deliveries per day, divide the number of deliveries per year by 365. If you have number of deliveries per month or in the last month, instead of number of deliveries per year, multiply the number of deliveries per month by 12 to get the number of deliveries per year. Once you have the number of deliveries per year you can calculate the number of deliveries per day. See Table 3 for an example of the calculations for Ethiopia. Note that

**Table 3. Sample selection based on number of deliveries per shift: Ethiopia example**

Location (Region)	Facility Name	Deliveries per Month (a)	Deliveries per Year (b = a*12)	Deliveries per Day (c = b/365)	Deliveries per Shift (d = c/3)	Cumulative Deliveries per Shift (e)	Cumulative Deliveries per Interviewer Day (f = e*2)	Cumulative Deliveries per Observation Period (g = f*2)	Adjusted Deliveries per Observation Period (h = f*3)
Addis Ababa	Yekatit Hospital	397	4,764	13.1	4.4	4.4	8.7	17.4	26.1
Addis Ababa	Gandhi Hospital	350	4,200	11.5	3.8	8.2	16.4	32.7	49.1
Oromiya	Adamma Zonal Hospital	300	3,600	9.9	3.3	11.5	22.9	45.9	68.8
Addis Ababa	Tikur Anbessa Hospital	250	3,000	8.2	2.7	14.2	28.4	56.9	85.3
Amhara	Gonder medical school hospital	240	2,880	7.9	2.6	16.8	33.7	67.4	101.1
SNNPR	Weliso hospital	240	2,880	7.9	2.6	19.5	38.9	77.9	116.8
Amhara	Felege Hiwot Hospital	210	2,520	6.9	2.3	21.8	43.6	87.1	130.7
Tigray	Mekelle Hospital	200	2,400	6.6	2.2	24.0	47.9	95.9	143.8
Hareri	Hiwot Fana Hospital	170	2,040	5.6	1.9	25.8	51.7	103.3	155.0
Hareri	Misrak Arbegnoch Hospital	160	1,920	5.3	1.8	27.6	55.2	110.3	165.5
Addis Ababa	Zewditu Hospital	158	1,896	5.2	1.7	29.3	58.6	117.3	175.9
Amhara	Debretabor Hospital	150	1,800	4.9	1.6	31.0	61.9	123.8	185.8
Amhara	debrermarkos Hospital	150	1,800	4.9	1.6	32.6	65.2	130.4	195.6
Addis Ababa	St. Pauls Hospital	147	1,764	4.8	1.6	34.2	68.4	136.9	205.3
Oromiya	Bishoftu District Hospital	140	1,680	4.6	1.5	35.7	71.5	143.0	214.5
Addis Ababa	Addis Ketema Health Center	136	1,632	4.5	1.5	37.2	74.5	149.0	223.4
Amhara	Dessie Hospital	120	1,440	3.9	1.3	38.6	77.1	154.2	231.3
Amhara	Debre birhan Hospital	120	1,440	3.9	1.3	39.9	79.7	159.5	239.2
Addis Ababa	Woreda 23 Health Center	92	1,104	3.0	1.0	40.9	81.8	163.5	245.3
Tigray	Adigrat Hospital	91	1,092	3.0	1.0	41.9	83.7	167.5	251.2
Tigray	Axum Hospital	90	1,080	3.0	1.0	42.9	85.7	171.4	257.2
Oromiya	Ambo Zonal Hospital	90	1,080	3.0	1.0	43.8	87.7	175.4	263.1
Tigray	Adwa Hospital	85	1,020	2.8	0.9	44.8	89.6	179.1	268.7
Addis Ababa	Woreda 24 Health Center	85	1,020	2.8	0.9	45.7	91.4	182.8	274.3
Addis Ababa	Woreda 17 Health Center	81	972	2.7	0.9	46.6	93.2	186.4	279.6
Tigray	Maichew Hospital	75	900	2.5	0.8	47.4	94.8	189.7	284.5
Oromiya	Shashemane Zonal Hospital	66	792	2.2	0.7	48.1	96.3	192.6	288.9
Amhara	Finote selam Hospital	60	720	2.0	0.7	48.8	97.6	195.2	292.8
Amhara	Weldiya Hospital	60	720	2.0	0.7	49.5	98.9	197.8	296.7

**Table 3. Sample selection based on number of deliveries per shift: Ethiopia example**

Location (Region)	Facility Name	Deliveries per Month (a)	Deliveries per Year (b = a*12)	Deliveries per Day (c = b/365)	Deliveries per Shift (d = c/3)	Cumulative Deliveries per Shift (e)	Cumulative Deliveries per Interviewer Day (f = e*2)	Cumulative Deliveries per Observation Period (g = f*2)	Adjusted Deliveries per Observation Period (h = f*3)
SNNPR	Attat Hospital	60	720	2.0	0.7	50.1	100.2	200.5	300.7
Addis Ababa	Ledeta Health Center	60	720	2.0	0.7	50.8	101.5	203.1	304.6
Dire Dawa	Dil Chora Hospital	60	720	2.0	0.7	51.4	102.9	205.7	308.6
SNNPR	Yirgalem Zonal Hospital	55	660	1.8	0.6	52.0	104.1	208.1	312.2
Addis Ababa	T/Haimanot Health Center	53	636	1.7	0.6	52.6	105.2	210.5	315.7
Oromiya	Asela Zonal Hospital	52	624	1.7	0.6	53.2	106.4	212.7	319.1
SNNPR	Arba Minch Hospital	52	624	1.7	0.6	53.8	107.5	215.0	322.5
Oromiya	Jimma Zonal Hospital	45	540	1.5	0.5	54.2	108.5	217.0	325.5
Oromiya	Neqamte Zonal Hospital	40	480	1.3	0.4	54.7	109.4	218.7	328.1
SNNPR	Hossana (Watchamo) Hospital	40	480	1.3	0.4	55.1	110.2	220.5	330.7
Oromiya	Dambidolo District Hospital	35	420	1.2	0.4	55.5	111.0	222.0	333.0
SNNPR	Butajera Hospital	35	420	1.2	0.4	55.9	111.8	223.6	335.3
Addis Ababa	Woreda 5 Kebele18 Public HC	33	396	1.1	0.4	56.3	112.5	225.0	337.5
Oromiya	Mettu Zonal Hospital	32	384	1.1	0.4	56.6	113.2	226.4	339.6
Tigray	Alamata Hospital	25	300	0.8	0.3	56.9	113.8	227.5	341.3
Oromiya	Deder District Hospital	25	300	0.8	0.3	57.2	114.3	228.6	342.9
Oromiya	Fichee Zonal Hospital	25	300	0.8	0.3	57.4	114.8	229.7	344.5
Oromiya	Chiroo Zonal Hospital	25	300	0.8	0.3	57.7	115.4	230.8	346.2
Oromiya	Gelemso Zonal Hospital	25	300	0.8	0.3	58.0	115.9	231.9	347.8
Tigray	Abiadi Hospital	24	288	0.8	0.3	58.2	116.5	232.9	349.4
Tigray	Kuha Hospital	24	288	0.8	0.3	58.5	117.0	234.0	351.0
Tigray	Wikro Hospital	23	276	0.8	0.3	58.8	117.5	235.0	352.5
Oromiya	Ginde Beret District Hospital	22	264	0.7	0.2	59.0	118.0	236.0	354.0
Amhara	Lalibela Hospital	20	240	0.7	0.2	59.2	118.4	236.8	355.3
Oromiya	Shambu District Hospital	18	216	0.6	0.2	59.4	118.8	237.6	356.4
Amhara	Borumada Hospital	15	180	0.5	0.2	59.6	119.1	238.3	357.4
Amhara	Mehal Meda Hospital	10	120	0.3	0.1	59.7	119.4	238.7	358.1

the Ethiopia example includes health centers. This sampling plan only refers to hospitals, but other facilities such as health centers could be included in the sample if they are expected to provide AMTSL and have sufficient deliveries for observations to occur.

The second step is to calculate the number of deliveries per work shift. We want at least one delivery per shift for a hospital to be included in the sample. Divide the number of deliveries per day by 3 to get the number of deliveries per shift. We assume that there are 3 shifts of 8 hours per day.

Step three is to sort the hospitals by number of deliveries per shift. These should be sorted from the highest to the lowest. The highest number of deliveries will be at the top of the shift and the lowest will be at the bottom.

We could stop here and include all facilities with more than one delivery per shift. Before doing that we want to know if such a selection will give us sufficient sample size based on observing deliveries 2 shifts per day for 2 days. If not, we can increase the number of days of observation. If it gives us too large a sample, we can reduce the number of observation days, or increase the threshold for selection from one delivery per shift to a higher number.

In step four you will calculate the cumulative number of deliveries per shift. To the right of the other columns, calculate the cumulative deliveries by adding the number of deliveries per shift for a facility to the cumulative for the facility in the prior row.

Step five is to calculate the cumulative number of deliveries per interviewer day. By interviewer day we mean 2 shifts. That is to say that within one day we will have interviewers observe for the first two shifts. In the initial set of surveys it was decided that it would not be reasonable to expect interviewers to work through the night on the third shift, which is why two shifts has become the standard. The number of deliveries per interviewer day is found by multiplying the cumulative number of deliveries per shift by 2.

The sixth step is to calculate the cumulative number of deliveries per observation period. In this case we assume the observation period will be two days. Calculating the number of deliveries per observation period done by multiplying the number of deliveries per interviewer day by 2.

The seventh step is to make a preliminary determination of which facilities will be in the sample. Look down the column with number of deliveries per shift to find the last facility with one or more delivery per shift. For Ethiopia it is Ambo Zonal Hospital. Then move along the row for this facility to the column for cumulative deliveries per observation period. In this case it is 175 deliveries, which is not sufficient for the sample size of 213 deliveries suggested above.

The final step is to make adjustments to the calculations and sample if necessary. If the number of number of deliveries per observation period is either too high or too low, we can make adjustments to the number of observation days or the threshold for number of

births per shift. In cases where there are more than double the number of deliveries required, you can reduce the number of observation days from two to one. Alternatively, you could increase the threshold for number of births per shift from one to two. This is less desirable because it will mean that facilities with fewer births will not be included in the sample, which will make the sample less representative. In cases where there are too few deliveries, increasing the number of observation days is a reasonable solution. Reducing the threshold for number of deliveries per shift is less reasonable, because many more of these facilities will be required to obtain a sufficient sample size, which would quickly increase the cost of the survey. Ethiopia is an example of there not being enough deliveries. The solution chosen was to increase the number of observation days from 2 to 3. When this is done there will be 263 deliveries. See the last column in Table 3.

We finish the process for Ethiopia by selecting all facilities with one or more deliveries per shift. See Table 4. One facility with less than one delivery per day was included in the sample to make the sample more representative of facilities *in different regions*. It is reasonable to do this, but it makes description of the sample less clean.

**Table 4. Selected facilities based on number of deliveries per shift: Ethiopia example**

Number	Location (Region)	Facility Name	Deliveries per Month	Deliveries per Shift	Deliveries per Observation Period
1	Addis Ababa	Yekatit Hospital	397	4.4	26.1
2	Addis Ababa	Gandhi Hospital	350	3.8	49.1
3	Oromiya	Adamma Zonal Hospital	300	3.3	68.8
4	Addis Ababa	Tikur Anbessa Hospital	250	2.7	85.3
5	Amhara	Gonder medical school hospital	240	2.6	101.1
6	SNNPR	Weliso hospital	240	2.6	116.8
7	Amhara	Felege Hiwot Hospital	210	2.3	130.7
8	Tigray	Mekelle Hospital	200	2.2	143.8
9	Hareri	Hiwot Fana Hospital	170	1.9	155.0
10	Hareri	Misrak Arbegnoch Hospital	160	1.8	165.5
11	Addis Ababa	Zewditu Hospital	158	1.7	175.9
12	Amhara	Debretabor Hospital	150	1.6	185.8
13	Amhara	debrermarkos Hospital	150	1.6	195.6
14	Addis Ababa	St. Pauls Hospital	147	1.6	205.3
15	Oromiya	Bishoftu District Hospital	140	1.5	214.5
16	Addis Ababa	Addis Ketema Health Center	136	1.5	223.4
17	Amhara	Dessie Hospital	120	1.3	231.3
18	Amhara	Debre birhan Hospital	120	1.3	239.2
19	Addis Ababa	Woreda 23 Health Center	92	1.0	245.3
20	Tigray	Adigrat Hospital	91	1.0	251.2
21	Tigray	Axum Hospital	90	1.0	257.2
22	Oromiya	Ambo Zonal Hospital	90	1.0	263.1
23	Dire Dawa	Dil Chora Hospital	60	0.7	267.0

## References

Casagrande, JT, MC Pike and PG Smith. (1978) An Improved Approximate Formula for Calculating Sample Size for Comparing Two Binomial Distributions. *Biometrics*, 34:483-486.

Fleiss, JL, B Levin and MC Paik. (2003) *Statistical Methods for Rates and Proportions*. New York: John Wiley & Sons.

Hulley, SB and SR Cummings. (1988) *Designing Clinical Research*, Williams & Wilkins.

Lemeshow, S, DW Hosmer Jr., J Klar, and SK Lwanga. (1990) *Adequacy of Sample Size in Health Studies*. New York. John Wiley & Sons.

## Appendix A: Sample Calculations

### Sample size

Sample size can be calculated directly from the equations if the AMTSL prevalence, degree of precision or level of change is not in the tables. The sample size equation for point estimates is (Lemeshow et al. 1990):

$$n = \frac{z^2 \cdot p \cdot q}{r \cdot e^2} Deff \quad \text{Eq. A1}$$

where  $n$  is the sample size,  $z$  the z-score,  $p$  the level of AMTSL prevalence,  $q$  the percentage that did not receive AMTSL ( $q = 1-p$ ),  $Deff$  the design effect,  $r$  the response rate, and  $e$  the precision or the distance from the prevalence estimate in either direction. In Table 1, the z-score is set at the 95% level for a two-tailed test ( $z=1.96$ ). The response rate was set to 90% (0.9 in the equation). The design effect in the table has been set to 2.0. A simple way to change the design effect is to divide the value in the table by 2.0 and multiply this by the appropriate design effect. Note that some surveys, like the DHS, report  $Deft$  which is the square root of the design effect. The calculations must be adjusted accordingly.

The sample size equation for measuring change over time is (Lemeshow et al. 1990; Fleiss, Levin and Paik 2003):

$$n = \frac{\left( z_{1-\alpha} \sqrt{2 \cdot p \cdot q} + z_{1-\beta} \sqrt{p_1 \cdot q_1 + p_2 \cdot q_2} \right)^2}{(p_2 - p_1)^2} \quad \text{Eq. A2}$$

where  $z_{1-\alpha}$  the z-score for the level of significance  $\alpha$ ,  $z_{1-\beta}$  the z-score for the power  $1-\beta$ ,  $p_1$  and  $p_2$  the levels of AMTSL prevalence at time 1 and 2, and,  $q_1$  and  $q_2$  the percentage who did not receive AMTSL at times 1 and 2. Estimate  $p$  with the average for the two time periods,  $p = (p_1 + p_2)/2$  and  $q = 1-p$ .

Along with adjustments for design effect and response rate, a standard adjustment is normally made to Equation A2 (Casagrande, Pike and Smith 1978).

$$n' = \left( \frac{n}{4} \right) \frac{\left( 1 + \sqrt{1 + \frac{4}{n|p_1 - p_2|}} \right)^2}{r} Deff \quad \text{Eq. A3}$$

Values in Table 2 are based on Equation A3. The level of significance  $\alpha$  in the table is set to 5% for a one-tailed test. The power  $1-\beta$  is set to 80%. The response rate is 90%.

### Sample weights

The goal of sample selection is to create a sample that accurately represents the entire population of interest. Aspects of the design and errors in the sampling process can

reduce this accuracy. Some sample designs over sample a subpopulation in order to ensure that subpopulation is sufficiently represented in the sample. Over sampling creates a bias in the analytical results based on the sample because of the over representation of a subpopulation. Errors in the sampling process can also bias the results. Weights are incorporated into the analysis to correct for such bias.

The sampling process outlined here is close to unbiased by design for the population of interest. In this case the population of interest is all facilities with more than one delivery per shift. There is no over or under representation of facilities in this sample, because all facilities in the population of interest are included in the sample.

Where the sample is not completely unbiased is in terms of deliveries. If we assume that the number of deliveries per day were constant in a facility, the sample would be unbiased. Given a fixed time of observation, the number of deliveries observed in a facility would be in proportion to the number of deliveries per year. If there were one full day of observation in a facility, the number of deliveries per year would equal 365 times the deliveries observed for the day. Deliveries, however, do not occur uniformly over time, which means the number of deliveries per day in a facility is not constant. Depending on the days when observation occurs, the number of deliveries could be higher or lower than the average number of deliveries per day. Based on the number of deliveries observed, some facilities will be overrepresented and some underrepresented in the sample.

Weights can be calculated to correct for the bias due to the number of deliveries observed not equaling the average number of deliveries. These weights are of the form:

$$\omega_i = \left( \frac{B_{Yi}}{B_{Oi}} \right) \left( \frac{D_o}{D_Y} \right) \quad \text{Eq. A4}$$

where  $\omega_i$  is weight for the  $i^{\text{th}}$  facility,  $B_{Yi}$  the number of deliveries per year in the  $i^{\text{th}}$  facility,  $B_{Oi}$  the number of deliveries observed in the facility  $i$ ,  $D_o$  the number of days for which observations take place, and  $D_Y$  the number of days per year, which will be set to 365. The number of deliveries should include still births as well as live births. In some countries the national health management information system only records the number of live births in health facilities. Live births can substitute for deliveries in those countries.

In some statistical packages, when the weights in Eq. A4 are applied, the total weighted sample size ( $n_\omega$ ) will be different than the actual sample size of deliveries observed ( $n_o$ ). A normalization factor can be created to make the weighted sample size equal the observed sample size. The formula for the normalization factor ( $f_{Ni}$ ) is given in Eq. A5.

$$f_{Ni} = \frac{n_o}{\sum \omega_i} \quad \text{Eq. A5}$$

The total weight ( $\omega_{Ti}$ ) is the normalization factor times the unnormalized weight.

$$\omega_{Ti} = f_{Ni} \omega_i \quad \text{Eq. A5}$$

Calculations for weights based on sample designs that randomly select facilities among those with sufficient deliveries would be more complex than those shown here. A sampling statistician would be required to design such a sample and calculate the weights.