DEPARTMENT OF BIOSTATISTICS
VANDERBILT UNIVERSITY

Survey Sample Size Calculations:

SCIP:Improving health and livelihoods of children, women and families in the Province of Zambézia, Republic of Mozambique

Biostatistician:
Meridith Blevins, M.S.

Faculty Statistician:
Bryan Shepherd, Ph.D.

June 13, 2010

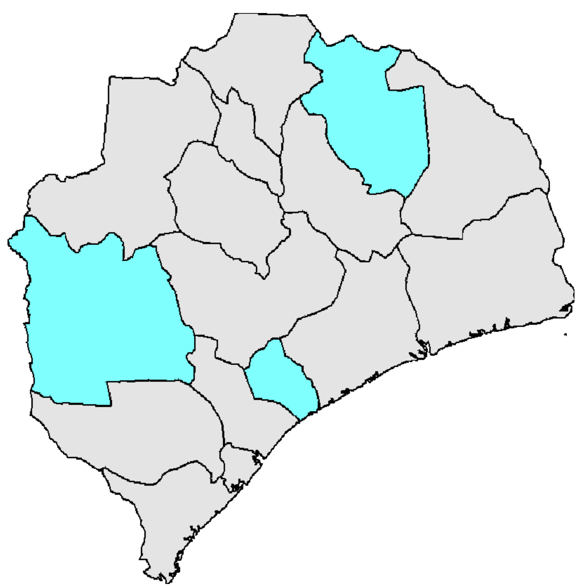
1 Purpose

We want to monitor the effectiveness of SCIP initiatives in Zambézia province through administration of a baseline and 5 year follow-up survey.

2 Sampling Frame

The sampling frame for the baseline survey will rely on the 2007 Mozambique Population and Housing Census. A list of *Enumeration Areas* (EA) covering the province will be made available with basic housing and population information and cartographic materials. The population for Zambézia is estimated at 3,794,489 among 918,025 households which are divided into 9,073 EA. There are 155,202 urban households (1,458 EA) and 762,823 rural households (7,615 EA).

The National Institute of Statistics (INE) for the Republic of Mozambique is charged with collecting census information of residents. In recent Province of Zambézia census 2007 results, there are 3,848,276 residents of which 548,466 (14%) reside in urban areas. The Province of Zambézia is divided into 17 districts: Cidade de Quelimane (193,343), Alto Molocue (271,650), Chinde (119,898), Gile (169,285), Gurue (297,935), Ile (289,542), Inhassunge (91,196), Lugela (135,485), Maganja da Costa (276,881), Milange (498,635), Mocuba (300,628), Mopeia (115,291), Morrumbala (358,915), Namacurra (186,410), Namarroi (125,999), Nicoadala (231,850), and Pebane (185,333). We intend to collect two representative samples. One sample of Zambézia province will allow for general estimates. A second concentrated sample in Alto Molocue, Morrumbala, and Namacurra will allow for precise estimates and smaller precision of the difference from baseline to five year. These three districts make up approximately 20% of the provincial population.



3 Two-Stage Cluster Sample Size

To monitor the effectiveness of SCIP initiatives while giving consideration for cost of administration, we will implement a two-stage cluster sampling design. Determining optimal sample size is critical because it requires a trade-off between the budget and the desired survey precision.

We will determine the number of clusters to be selected and the number of individuals to be interviewed in each EA in order to achieve the desired precision within the survey budget[1]. The number of individuals interviewed per EA is set at 15 to 17. ICC measures the similarity of the individuals on one survey characteristic within a cluster, δ.

Next we calculate the design effect which quantifies the increase in the standard error of the estimate due to the sampling procedure used[2,3]. The design effect increases with cluster size $\bar{n}$ and decreases for small ICC, $D = 1 + \delta(\bar{n} - 1)$. Once the optimal cluster size $\bar{n}$ is obtained, the number $\bar{m}$ clusters to be selected within strata may be determined by

$$\bar{m} = \frac{p(1-p)D}{s^2\bar{n}} \tag{1}$$

For a confidence interval of $\pm 5\%$ we shall need $s = 0.025$. If we have some idea of the proportion p in advance, then it may be used. The most conservative approach is to use $p = 0.5$ as a guess since this maximizes the standard error, s.

Table 1 gives sample size estimates for surveys of 2 strata. The proportion of interest is set at $p = 0.5$. Precision varies to demonstrate the different estimates of number of interviews ($\bar{n}$) and number of EA ($\bar{m}$). Based on average intracluster correlations for selected indicators averaged over 48 different DHS surveys[1], we set δ=0.15.

Table 1: Sample Sizes

Precision (%)	$\bar{n}$	$\bar{m}$	Per stratum	2 strata[†]	Precision (%) 20% respondents[§]	Precision (%) 50% respondents[§]
3.5	15	169	2535	5070	8	5
4	15	130	1950	3900	9	6
4.5	15	103	1545	3090	10	6
5	15	83	1245	2490	11	7
10	15	21	315	630	22	14
3.5	16	166	2656	5312	8	5
4	16	127	2032	4064	9	6
4.5	16	101	1616	3232	10	6
5	16	82	1312	2624	11	7
10	16	21	336	672	22	14
3.5	17	164	2788	5576	8	5
4	17	125	2125	4250	9	6
4.5	17	99	1683	3366	10	6
5	17	80	1360	2720	11	7
10	17	20	340	680	22	14

[†] 2 strata correspond to differing levels of initiatives in Zambézia and within these, there will be 2 strata (e.g., rural/urban or near/far from medical center).

[§] If the desired survey characteristic may only be measured for a smaller percentage of households, then the precision of this estimate drops (e.g., if 20% of households have children under 5 or 50% of households have a female respondent aged 15-64).

4 Sample Size

The total sample size is 3960 households.

4.1 Alto Molocue, Morrumbala, and Namacurra

The enumeration areas of three Zambézian districts are first stratified by two intervention levels (0, 1 or more) with 1545 households in each group. The sample must consist of no fewer than 206 EA. Each intervention level has 103 EA with 15 interviews each. Sample size is motivated by 4.5% precision within each of 2 strata.

[1] Aliago A, Ruilin R. Optimal Sample Sizes for Two-Stage Cluster Sampling in Demographic and Health Surveys *Demographic and Health Research* 30(2006).

[2] Cochran WG. *Sampling Techniques.* New York, NY: Wiley; 1977.

[3] Bennet S, Woods T, Liyanage WM, Smith DL. A Simplified General Method for Cluster-Sample Surveys of Health in Developing Countries *World Health Statistics Quarterly* 44(1991).

Probabilistic selection criteria was used throughout the 3 district sample, except at the level of the household, when random sampling will be used. To ensure a representative sample of households with access to health care, we used implicit stratification during selection of EA. Provincial districts have an average of 7 health centers and health posts and another 7 community posts[4]. There are, however, differences across districts in the number of community posts, with some districts having many and many districts having none. Ministry of Health centers are known to be located in the center (*Sede*, first locality) of each Administrative Post at a minimum. This stratification is used instead of the more traditional urban/rural strata. Because the great majority of the population in Zambézia lives in rural areas, and because access to health care and other services is a key factor in the OGUMANIHA project, our sample will contain approximately 30% of these areas with known health centers.

4.2 Zambézia-wide

To maintain some degree of provincial generalizability, a sample of households will be selected for interview throughout the remaining province. Because 20% of the population is well represented in our concentrated sample of 3 districts (above), we will need fewer households from the remaining 80% of Zambézia. Subsetting to the remaining 14 districts, we will sample 58 EA with 15 interviews each, that is 870 households. Sample size is motivated by 15 interviews per EA and it yields slightly better than 5% precision. Probabilistic selection criteria will be used throughout the Zambézia-wide sample, except at the level of the household, when random sampling will be used.

In order to first provide a report of Zambézia-wide estimates with reasonable baseline precision (=6.5%), a random selection of 9 EA (with 15 interviews each) from the 3 district allocation will need to be collected in addition to the 58 EA from the Zambézia-wide allocation.

5 Estimating the change from Baseline to 5 year follow-up

If the purpose is to estimate the change in proportion from Baseline to 5 year follow-up, then the standard error of the change will need to be estimated. This is larger than the standard error of the 5 year point estimate, because of the imprecision of the estimate of the proportion from the baseline survey. If the desired estimate is a difference in proportions from one survey to the next and the desired precision remains the same, then the precision of the difference is slightly larger than the precision at baseline or 5 years.

Table 2: Sample Precision

Precision within strata†	0.035	0.040	0.045	0.050	0.100
Precision of the difference baseline to 5 year for independent samples	0.049	0.057	0.064	0.071	0.141
Precision of the difference baseline to 5 year for same clusters	0.041	0.047	0.053	0.059	0.118
Precision for Alto Molocue, Morrumbala, and Namacurra (2 strata)	0.025	0.028	0.032	0.035	0.071
Precision of the difference baseline to 5 year for independent samples	0.035	0.04	0.045	0.05	0.1
Precision of the difference for binary intervention effect for independent samples	0.07	0.08	0.09	0.1	0.2
Precision of the difference baseline to 5 year for same clusters	0.029	0.033	0.038	0.042	0.084
Precision of the difference for binary intervention effect for same clusters	0.059	0.067	0.075	0.084	0.167

† 2 strata correspond to differing levels of interventions.

6 Baseline Survey Anthropometrics

6.1 Overview[5]

We wish to assess nutritional status of Zambézian children under age five on the basis of anthropometric indicators. These are based on physical body measurements such as height or weight along with attributes

[4]Lindelow M, Ward P, Zorzi N. Primary Health Care in Mozambique: Service Delivery in a Complex Hierarchy. *Africa Region Human Development Working Paper Series.* The World Bank, 2004.

such as age and sex.

Weight-for-height (W/H)

W/H measures body weight relative to height. Weight-for-height is normally used as an indicator of current nutritional status, and can be useful for screening children at risk. Extreme cases of low W/H are commonly referred to as "wasting".

Height-for-age (H/A)

H/A reflects cumulative linear growth. H/A deficits indicate past or chronic inadequacies nutrition and/or chronic or frequent illness. Extreme cases of low H/A, where shortness is interpreted as pathological, is referred to as "stunting".

Weight-for-age (W/A)

W/A reflects body mass relative to age. W/A is a composite measure of height-for-age and weight-for-height, making interpretation difficult. The term "underweight" is commonly used to refer to severe or pathological deficits in W/A. W/A is commonly used for monitoring growth and to assess changes in the magnitude of malnutrition over time.

Z-score (standard deviation score)

Anthropometric indices are constructed by comparing relevant measures with those of comparable individuals (in terms of age and sex) in the reference data (WHO child growth charts). We express this comparison as a Z-score (standard deviation score): the difference between the value for an individual and the median value of the reference population for the same age or height, divided by the standard deviation of the reference population.

6.2 Prior knowledge

A 2008 survey from USAID indicated that median z-scores fell between -1 and -1.99 for H/A and W/A and above -1 for W/H among eight Zambézian districts in the MYAP program. Previous DHS surveys give mean estimates of z-scores:

Table 3: Prior Estimates of Mozambican Anthropometrics in Children 0-59 Months

	Mozambican z-score (sd) 1996/1997	Rural Mozambican z-score (sd) 2003
Weight-for-height	-0.15 (1.34)	-0.12 (not given)
Height-for-age	-1.88 (1.74)	-1.84 (1.5)
Weight-for-age	-1.28 (1.31)	-1.24 (not given)

The 2007 census counted 826,619 young children among 3,848,276 persons yielding 21.5% aged 0-59 months. Another publication by a hunger task force identified 16.8% of the population as aged 0-59 months. A third publication for a 2008 USAID baseline survey identified 77% of Zambézian households with one or more child aged 0-59 months.

[5] O'Donnell O, van Doorslaer E, Wagstaff A, and Lindelow M. *Analyzing Health Equity Using Household Survey Data.* The World Bank, 2008.

6.3 Sample Size

To determine the number of children 0-59 months needed for anthropometrics, we need to select a minimum detectable difference. Because very few children fall outside of normal range for W/H, we will only focus on improvements in H/A and W/A. We will assume equal variance from baseline to five year sample. We want to be able to reject the null hypothesis that the anthropometrics are unchanged from baseline to 5 years with probability (power) 0.90 and have the Type I error probability associated with our test of this null hypothesis equal to 0.05. Note, the anthropometrics design effect is smaller than the overall survey design effect because previous DHS surveys[1] indicate low intracluster correlation, $\delta = 0.07$.

$$N = \frac{(1 + \delta(\bar{n} - 1))(Z_\alpha + Z_\beta)^2(2\sigma^2)}{(\bar{X}_2 - \bar{X}_1)^2} \quad (2)$$

To detect an improvement in mean z-score for H/A from -1.8 to -1.4, we will need to measure a minimum of 482 children 0-59 months with 10 HH per EA. To detect an improvement in mean z-score for W/A from -1.25 to -1 we will need to measure a minimum of 941 children 0-59 months with 10 HH per EA.

6.4 Sample Allocation

If we rely on the prior USAID survey and expect 70% of Zambézian households to have at least one child 0-59 months, then we would plan to measure children as follows:

[1] 3 districts: 37 EA will yield approximately 10 HH with a child 0-59 months.

[2] Zambezia-wide: All 58 EA will yield approximately 10 HH with a child 0-59 months.

In total, approximately 950 children under five should be considered for anthropometrics. It is permissible, though not necessary for multiple children to be measured per household.