

World Bank/ MPND/Stats

Vulnerability and poverty assessment survey – 2004

---

Sampling design

Shyam Upadhyaya

## **1. Background**

The Vulnerability and poverty assessment survey (VPA) was conducted in 1997/98 to collect the wide range of data measure the poverty, deprivation and vulnerability of population arising from geographical, social and economic conditions of Maldives. The survey was the most comprehensive statistical investigation in terms of its geographical coverage and statistical data items. Major findings and results of the survey were presented in a report that provided among others a composite index of human vulnerability (HVI) at the national, atoll and island level. Results of the survey helped government formulate policy of regional development. Unlike for other countries of the sub-continent, income poverty was a matter of lesser importance for Maldives than its vulnerability caused mainly by geographical isolation of islands, where opportunities of population for job as well as for essential public services are severely limited.

Results of VPA-97 provided important information for formulation of development strategy in the past years. Much have changed in Maldives since 1998. The country enjoyed higher economic growth, tourism industry was flourishing all these years and construction sector boomed. How far economic development has been successful to change the living condition of Maldivian population at large? Answer to this question can be obtained from a statistical survey that is compatible in terms of methodology and coverage to that of VPA-97. The Ministry of Planning and National Development has decided to conduct a follow up survey this year with technical assistance of UNDP and the World Bank.

## **2. Objectives of the survey**

VPA-04 will be conducted with the main objective of producing wide range of statistics on various aspects of poverty and vulnerability of households. The survey will provide the comparative data with VPA-97 that facilitates the analysis of social and economic impact of recent development on the living conditions of population. The survey results will allow to measure the changes occurred in individual islands, in atolls and in country in general since the last survey.

The survey will be conducted in close cooperation of international experts and national staffs of the Statistics section of MPND. Since 1997 MPND/Stats has been successfully conducting the series of statistical surveys to improve the national database in different areas. The co-operation under VPA will further enhance its technical capability of undertaking complex statistical operations.

The VPA questionnaire comprises different components. Being the largest survey in terms of its geographical coverage, it will produce a new area frame with the recent number of households, labour force statistics, household income and expenditure and other information thereby update the current national database.

### 3. Sampling strategy

VPA-97/98 report does not provide the information about the variance but an approximation can be made for the standard error of estimate of one of the key variables. The main indicator of VPA was the Human vulnerability index calculated from the poverty incidence indices at different income levels: such as the percentage of population with less than MRF 7.5, MRF 10 and MRF15. For the sampling design purpose we take the Poverty index estimated from the sample data ( $p$ ) for MRF 10 which was computed as the proportion of the population with the per-capita income less than MRF 10 per day to the total population. The standard error of this estimate  $se(p)$  is given by:

$$se(p) = \sqrt{deff \cdot (1-f) \frac{p(100-p)}{n-1}} \quad (1)$$

where:

$se(p)$  – standard error of the estimated poverty index from the sample data  
 $deff$  – design effect, inflation of variance due to stratification, unequal probability sampling and clustering in complex designs and

$(1-f)$  – factor of finite population correction  $f = \frac{n}{N}$

Estimated standard error of the VPA for the poverty index is given in Table 1 assuming that the design effect was 2. For the national level estimate of  $p$  we can define the 95% confidence interval between 19.9 to 24.1 [ $22 \pm t.se$ ; where  $t = 1.96$  for the precision level of 95%].

**Table 1:**  
**Estimated poverty index and its standard error for VPA-97**

	$p$ in %	n	1-f	$se(p)$ in %
Maldives	22	2778	0.919	1.07
Male'	10	300	0.954	2.40
Atolls	26	2478	0.910	1.19

The above formulae is not the most appropriate one to compute the standard error for the systematic selection method used in VPA. Better approximation of the standard error for the survey of VPA type can be obtained from linearisation using Taylor series expansion or replication methods. However, at this stage computation is made to examine whether the same sample size can be adopted for VPA-04. The level of precision of VPA-97 results shows that the same sample size can be retained also for VPA-04.

### 4. Sampling in atolls

As previous, sampling of islands is not considered appropriate, because the level of vulnerability is very much determined by the local conditions. The survey will cover all 200 inhabited islands. These islands are different in population sizes. A minimum sample of 10 households will be allocated for each island with less than 1500 inhabitants or approximately 200 households (average household size in atolls is around 7 persons according to the last population census). For larger islands, sampling rate will be

increased by 10 households for every 1500 inhabitants. Such distribution fairly satisfies the proportional allocation scheme thereby reduces the variance of results arising from disproportionate allocation. The total number of households to be sampled from all atolls will be 2480 households.

**Table 2:  
Size and allocation of samples in atolls**

Size-group	Number of islands	Population	Number of households	Sampling rate per island	Number of households to be sampled
1500 and less	168	103116	19446	10	1680
1501-3000	26	49748	6740	20	520
3001-4500	2	6919	1053	30	60
4501-6000	1	4893	179	40	40
6001-7500	1	6581	1018	50	50
7501-9000	1	7528	1251	60	60
More than 9000	1	9461	1408	70	70
<b>Grand Total</b>	<b>200</b>	<b>188246</b>	<b>31095</b>		<b>2480</b>

Source: Population and household data taken from the Population census 2000, MPND

In this sampling plan each island virtually becomes an independent stratum. Selection of households will be carried out within each island independently from others. Therefore, it facilitates aggregating island data by different grouping relevant to statistical analysis. Number of households to be sampled for each island are given below.

**Table 3:  
Number of households to be sampled by islands**

Name of islands	Number sample households per island
Hithadhoo	70
Fuvammulah	60
Kulhudhuffushi	50
Thinadhoo	40
Naifaru Hinnavaru	30
Feydhoo, Dhidhdhoo, Kadholhudhoo, Eydhafushi, Viligili, Gamu, Hoarafushi, Maradhoo, Ihavandhoo, Thulhaadhoo, Velidhoo, Maafushi, Fonadhoo, Alifushi, Mahibadhoo, Gadhdhoo, Meedhoo, Maakadoodhoo, Kaashidhoo, Holhudhoo, Nolvivaramu, Maamigili, Thimarafushi, Maduvvari, Komandoo, Nilandhoo	20
All other inhabited islands	10

## Partial overlapping sample

In order to ensure the data comparability of two surveys half of the samples in all islands will be retained from those selected for VPA-97. Partial overlapping of samples for successive surveys has certain advantages. A completely repeated panel can give the information about the changes of variables of interest, but ignores the effect of changes outside the panel. In contrary, an independent sample in the successive period cannot measure the changes occurred in individual units. Partial overlap balances the advantage and disadvantages of both methods.

There are also certain gains in reduction of variance by using the same sampling units in the successive survey. Suppose, we are conducting two surveys in different time periods and the variable to be estimated is  $p$ , say it denotes the proportion of population living below the poverty line. The variance of the estimated change of the difference of  $s_d^2 = p_1 - p_2$  is given by:

$$\begin{aligned} s_d^2 &= s_{p_1}^2 + s_{p_2}^2 - 2s_{p_1 p_2} \\ &= s_{p_1}^2 + s_{p_2}^2 - 2\rho s_{p_1} s_{p_2} \end{aligned} \quad (2)$$

where,  $s_p^2$  is the variance of estimates, suffix 1, 2 stands for period and  $s_{p_1 p_2}^2$  denotes the covariance and  $\rho$  - coefficient of correlation.

When the estimated proportion does not change sharply we can assume that variance of estimates of two periods are approximately equal (for example, if the poverty index falls to 17% from the earlier rate of 22% its variance will change merely by 0.08. Therefore,  $s_{p_1}^2 \approx s_{p_2}^2$ . Then variance of the difference would appear as,

$$s_d^2 = 2s_p^2 - 2\rho s_p^2 = 2s_p^2(1 - \rho) \quad (3)$$

When the same sample of households are taken, survey data are highly correlated thus correlation coefficient  $\rho$  reaches up to 0.8. In this case, variance of difference will decrease significantly. If we take the same clusters (in our case, islands) but different households, the value of  $\rho$  will be much smaller around 0.25. In case of completely new sample there would be no correlation i.e.  $\rho = 0$ , so higher the variance of difference.

To measure the gain of a partial overlap of the sample by reducing the variance of difference, we multiply the correlation coefficient  $\rho$  by a factor  $F$  that equal to the proportion of overlap. So the variance of difference would be :

$$s_d^2 = 2s_p^2(1 - F\rho) \quad (4)$$

It means with the value of  $\rho=0.8$  and  $F= 0.5$  (proportion of overlapping sample) the variance of difference will be less by 40%. Practical implication of above remarks is that not all the ten households but only 5 new households will be selected for each island with the population of 1500 and less and subsequently half of the sample of those given in Table 3 for other islands.

When half of the sample is overlapping, there is still a high degree of correlation between the samples of two periods. Thus the “old sample” still holds the influence on major characteristics to make the data set highly comparable for the growth measurement.

### **Selection procedure**

Information available about the households refer to the Population census of 2000. Therefore, it is necessary to have a fresh listing of households. Listing of household should be carried out in a systematic manner choosing a direction how the enumerators would move in the listing process. Normally a route (clock-wise or anti-clock-wise) of listing should be fixed. MPND/Stats has the good experience of listing households. The important thing to note that the households will be selected systematically with random start and this method gives better results if the listing is made in an order. Samples taken from the list arranged in order creates implicit strata of each interval. Systematic selection is simple, especially when the total number of units  $N$  is an integral multiple of the desired sample size  $n$ . Then an interval is calculated as

$k = \frac{N}{n}$  and the random start is made between 1 to  $k$ . If the  $N$  is not an integral multiple then chose  $k$  so that  $N$  is greater than  $nk$ .

Let us take an example of Finey island, which had 71 households in the Census 2000. Suppose, we get 74 households from the listing this time. First, we identify 10 households selected in VPA-97 and select 5 of them at random. From remaining 64 households we select 5 households systematically. Because, 64 cannot be divided by 5 we can take  $k=12$  so that  $N > nk$  or  $64 > 60$ . So we take the random start between 1 to 12 and select every twelfth household into sample.

### **If a household selected in VPA-97 does not exist any more**

Some households of the panel from the sample of VPA-97 may not exist any more in the island. First, households in the panel of the VPA-97 should be identified in the new list. If all households are found then sampling procedure may begin. If there are the cases when “old household” could not be found we have to apply different rules of replacement.

1. If the old household has moved away from the island then we consider it as a loss of panel household, thus the number of households in panel will decrease. We take the sample of 5 from the remaining “old households”. However, if in place (dwelling) of the “old household” we find the new household from the same family we regard it as a match case and consider it as an “old household”.
2. If we have the match of the household in our new list but the dwelling unit is different, we regard it as a household of the panel. It can happen when the household has moved to another place in the same island. Similar situation may also arise if the dwelling unit has been demolished. Again we try to find the household in the new list. If it could not be identified, we again consider it as a loss of panel household and follow the rule (1).

3. It is very unlikely that more than half of the households from the earlier sample does not exist. If so happens, number of new samples should be increased so that the total number of households selected from an island is 10 throughout all islands.
4. We assume that the household once identified in the listing will be available for interview. Non-response rate in the household surveys is quite negligible in Maldives, especially in atolls. In case the response from a selected household could not be attained (nobody at home or temporarily not at home, due to family vacation, or any emergency) substitution of sample households is allowed. Such substitution should be made from the respective panels, which means that the “old household” can be replaced from the panel of VPA-97, and new household from the rest.

If the households records of VPA-97 are readily available, it would also be advisable to carry out the listing in the same order (same route). In that case, both the selection from earlier sample as well as new sample should be made systematically, which would create a pair within each implicit stratum. Such arrangement greatly facilitates the estimation of sampling error using replication or interpenetrating sub-sample methods.

## 5. Sampling in Male’

Sampling in Male’ will be different for many reasons from those in atolls. First of all, there will be no panel in Male’ and a completely new set of samples will be taken. Second, in order to avoid the listing of all households, a two-stage self-weighting design will be applied. Male will be stratified by 5 wards and selection will be made within each ward. At the first stage, enumeration blocks will be selected probability proportion to the size (PPS) of blocks in terms of the number of households and at the second stage a fixed number of 10 households will be selected using systematic sampling from each selected block. In such case, block will be a primary sampling unit (PSU) and the household – the secondary sampling unit or elements.

Selection probability of a block for PPS selection equals  $f_1 = a \cdot \frac{m_j}{\sum m_j}$  where,  $a$

denotes the number of blocks selected and  $m_j$  – number of households in selected  $j$ -th cluster. Similarly, selection equation of a household is:  $f_2 = \frac{b}{m_j}$  where,  $b$  denotes the

number of households to be selected in a PSU. Then overall selection rate within the stratum is given by:

$$f_0 = f_1 \times f_2 = \frac{a \cdot b}{\sum m_j} \rightarrow const. \quad (5)$$

The first stage selection is probability proportional to the size and second stage selection is inversely proportional to the size of PSU. Such sampling plan results in a self-weighting design, where each household within the stratum has an equal probability of being selected. The main advantage of this sampling plan is that the mean, ratio and proportion from the sample can be used without weighting.

**Table 3:  
Size and allocation of sample in Male**

Male'	Population	Number of households	Number of blocks in total	Sample	
				Number of blocks	Number of households
Henveyru	18100	2488	76	8	80
Galolhu	13878	1813	59	6	60
Machchangolhi	13589	1748	57	6	60
Maafannu	22372	2928	108	10	100
Viligili	4291	601	17	2	20
<b>Total</b>	<b>72230</b>	<b>9578</b>	<b>317</b>	<b>32</b>	<b>320</b>

Source: Population and household data taken from the Population census 2000, MPND

The list of sample for different wards of Male' is given in Annex-1.

### Annex 1: List of selected blocks in Male by wards

	SN	Atoll/island code	Enumeration block no.	Number of households in the frame
<b>Henveiru</b>	1	1001	690	56
	2	1001	320	52
	3	1001	591	50
	4	1001	460	48
	5	1001	380	45
	6	1001	340	38
	7	1001	600	34
	8	1001	670	24
<b>Galolhu</b>	1	1002	440	47
	2	1002	60	48
	3	1002	220	28
	4	1002	40	44
	5	1002	160	34
	6	1002	130	43
<b>Machchangolhi</b>	1	1003	200	36
	2	1003	250	32
	3	1003	280	40
	4	1003	70	31
	5	1003	320	29
	6	1003	080	43
<b>Maafannu</b>	1	1004	650	25
	2	1004	820	40
	3	1004	740	27
	4	1004	280	25
	5	1004	331	21
	6	1004	640	46
	7	1004	352	21
	8	1004	310	55
	9	1004	770	25
	10	1004	780	42
<b>Villigili</b>	1	1005	140	54
	2	1005	160	52

## Annex 2: Substitution procedure

First, let us make clear that substitution is not recommended for non-response. Because the major variables reflecting the level of living of the non-responding household can be quite different from the one in the substitution list. From the past experience of household surveys, significant non-response is not expected in this survey too. However, due the small sample size at the level of islands, substitution is allowed in VPA in certain situations such as, family emergency, death of a household member or relatives, family vacation, prolonged absence of household (temporarily not at home).

Households on the island will be selected systematically from the list. Systematic sample creates an interval from which one sample is taken. In the example below, there were 41 households listed in an island, from which 10 households from VPA-97 were identified and separated. We divide households in either side into 5 groups, which is otherwise called as an implicit stratum.

New sample households			Panel households from VPA-97			
Implicit stratum	HH no. in sample	Selection process	Implicit stratum	HH no. in sample		
1	1		1	1		
	2			2		
	3		2	3	Failure	
	4	Sample		4	Substitution	
	5		3	5		
2	6		4	6		
	7			7		
	8		5	8		
	9	Sample		9		
	10		10			
3	11					
	12					
	13					
	14	Sample				
	15					
4	16		Suppose, the sample household of this group (19 <sup>th</sup> ) could not be surveyed. It can be substituted by one of the sampled household of this interval from 16 <sup>th</sup> to 20 <sup>th</sup> household. Say, randomly selected substitution is 17 <sup>th</sup> household.			
	17	Substitute				
	18					
	19	Sample				Failure
	20					
5	21					
	22					
	23	Sample				
	24					
	25					
	26					

Substitution of unattained household should be made by the household from the same implicit stratum. In the above example, 19<sup>th</sup> household in the sample could not be attained. This household can be substituted only by one of the randomly selected households between 16<sup>th</sup> to 20<sup>th</sup> household. For the panel households, each group always has 2 households. Failure of observation one of those requires that another household of the same group is taken into sample. If it were not possible, substitution can also be a household from the closest group.