
Mozambique Expenditure Tracking and Service Delivery Survey in the Health Sector

Sampling Design

Juan Muñoz¹, Consultant
April 2002

Introduction

I visited Maputo, Mozambique, from April 22 to 27, 2002, to help in the sampling design of the Expenditure Tracking and Service Delivery Survey (ETSDS) in the Health Sector. Oxford Policy Management (OPM) is implementing the survey in collaboration with Austral Consultoria, as a component of a DfID-financed project of support to the Mozambique Ministry of Planning and Finances.

The survey considers various units of observation: primary health facilities, facility staff members, clients and higher units in the health system such as the district and provincial administration.

My work included:

- A discussion of the sampling scheme for the facilities and other units;
- An assessment of the available lists of health facilities, with a view to the development of an adequate sample frame;
- The selection of the sample of districts to be visited by the survey
- The definition of the sampling procedures for choosing the facilities to be visited in the selected districts, and the staff members and clients to be interviewed in these facilities;
- The definition of the raising factors required to obtain unbiased estimates from the samples of facilities and other units of observation;

I worked in close collaboration with OPM's senior economist Mr Magnus Lindelöw, who was at the time of my mission pilot-testing the questionnaires to be used in the study, and I joined him in one of his field trips to assess the conditions faced by the part of the sampling process that needs to be implemented as a part of the survey field operation.

¹ Juan Muñoz. Sistemas Integrales. Casilla 13168 – Santiago – Chile. Tel: (56-2) 638-1841. Fax (56-2) 639-4554. Email: juan.munoz@ariel.cl

The rest of this document reports on the tasks conducted and gives recommendations for the implementation of the ETSDS sample and other survey activities. The forms developed for the selection of facilities, staff members and clients are reproduced as annexes.

The sample of facilities

The most natural starting point for the various samples required by the ETSDS is the sample of health facilities, since from both the analytical and operational standpoints it is important that the health staff and the clients are selected from the same establishments.

It is also clear that the sample of primary health care facilities needs to be selected in two stages: a certain number of districts first and then a certain number of health facilities in each of the selected districts. This strategy is also justified by analytical and operational reasons. On the one hand, it is important to observe more than a single health facility in each of the selected districts, to assess to what extent their performance can be explained in terms of local management parameters rather than by overall conditions of a higher level. On the other hand, the costs and managerial difficulties of the survey field operation are facilitated if the health facilities visited by the survey are clustered into some districts rather than scattered more sparsely throughout the territory.

An additional reason for implementing a two-stage sampling design in this case is the unavailability of a reliable sample frame – a list of all the health facilities currently operating in the country. Various versions of that list do exist, and indeed were used as a part of this mission to build what is possibly the most accurate roster now available, but it is likely that some facilities may be missing from the list, whereas others may appear twice with different names, or even that some facilities on the list are not actually operational.

The consolidated file prepared during this mission can be used, however, as a sample frame to select *the districts* to be visited by the survey (the first sampling stage). The list of the facilities actually operating in those districts can be built as a part of the survey field operation, and so can be implemented the selection of the health facilities to be visited in each district (the second sampling stage).

The sample frame contains 1,246 health facilities, distributed by region, level, and the urban / rural character of the district² they are located as shown in Table 1:

² The capital districts of each of the 11 provinces, along with the districts of Nacala-Porto, in the Nampula province, and Maxixe, in the Inhambane province (13 districts in total) were considered as **urban** for sampling purposes. The remaining 129 districts of Mozambique were considered as **rural**. This does not preclude, of course, the use of alternative classification criteria at the analytical stage of the study.

Table 1. Distribution of the Mozambique Health Facilities by Region, Level and Location

Region	First level facilities		Hospitals	Total
	Urban	Rural		
Norte	34	359	13	406
Zambezia and Tete	8	175	5	188
Center	40	284	10	334
South (exc. Maputo)	32	236	10	278
Maputo	36	-	4	40
Total	150	1,054	42	1,246

Within the constrain of the total sample size of 70 to 100 first level facilities that is envisioned, the only stratification criterion that seems necessary for the analytical purposes of the study is the distinction between Urban and Rural locations, since a non-stratified sample would contain too few facilities for any significant analyses of the urban areas separately. I therefore recommend that in the first stage all 13 urban districts are selected with certainty, along with a sample of 13 to 20 rural districts, selected with probability proportional to the number of first level facilities currently recorded in the available sample frame. The sample of rural districts should be implicitly stratified, to ensure that it is geographically distributed in proportion to the number of facilities in each region.

In the second sampling stage, I recommend to select three first level facilities with equal probability in each of the selected districts. This sampling scheme will provide a sample that is approximately self-weighted in the rural areas – in other words, it gives to all rural facilities in the country the same chance of being observed by the study.

The sample frame is contained in a workbook called “Frame.xls”, which should be considered as an integral part of this report. One spreadsheet in the workbook contains the formulae required to select the sample of rural districts once its final size is determined. Annex I explain the contents of the workbook and the procedures of sampling selection in detail.³

As said before, it is better to postpone the selection of the three first level facilities in each district until the operational phase of the survey – when the actual list of establishments actually operating in the district can be defined reliably – rather than doing it beforehand from the sample frame that is currently available. One of the forms reproduced in Annex II can be used to prepare the list of first level facilities and select from it a simple random sample of three establishments.

The sample of health staff

It is envisioned to interview three staff members in each of the health facilities visited by the survey. Depending on the number of facilities visited, the total size of this sample will thus be between 200 and 300. Since at the present moment there are no compelling

³ Workbook “Frame.xls” can be obtained from the author by simple request to juan.munoz@ariel.cl.

analytical reasons to stratify that sample by any criterion (such as the title, gender or level of the staff), or to exclude from the sample some categories of staff (such as secretaries or servants), it is recommended that the three members be selected by simple random sampling from the roster of all staff members actually working in the facility at the time of the survey. One of the forms in Annex II can be used to these effects.

The sample of clients

The idea is to select a random sample of all the clients visiting the facility on the day of the survey for non-urgent reasons. The total number of interviews that can be conducted in a day is a function of the questionnaire to be applied, which is still in its pilot-testing phase. It is estimated, however, that it will be possible to interview from 5 to 10 clients in each facility, for a total sample size of 500 to 1,000 clients.

The actual selection of the clients to be interviewed poses particular problems, because it cannot be left to the discretion of the interviewers, the staff of the health facilities, or the clients themselves, lest serious selection biases be introduced in the sample. Selection biases can also stem from informalities such as restricting the interviews to some particular hours of the day. The following procedure is recommended:

- Reach the health facility the day before, to agree with the director and the rest of the staff upon the method that will be applied. As a part of the discussion, estimate the number of clients expected the next day and determine the *sampling step* that will be used to select the clients to be interviewed (for instance, if 40 clients are expected and 8 interviews are to be conducted, the sampling step will be $40 / 8 = 5$, meaning that one out of each five clients will be interviewed).
- On the day of the survey, give to each client arriving to the facility a numbered card that s/he will have to show to the health staff in order to be served. The numbered card will be given to the client by the survey interviewer, who will ask the client to return the card to him/her before leaving, in order to receive a present.
- As each outgoing client returns the card and retrieves the present, the interviewer will use a standardized method to decide if the client needs to be interviewed.

The operational aspects of this procedure are formalized in one of the forms reproduced in Annex II. There are many aspects of the process that deserve comment:

- The brief field-testing conducted as a part of this mission suggests that the health staff can estimate the number of non-urgent clients expected the next day with reasonable accuracy, even if that number may depend on the weather and on factors such the day of the week (there are more clients on Mondays, for instance). Taking these factors into account as a part of the sample design would be quite complicated, but it would be useful to record them properly in the survey instruments, in order to use them, if necessary, as explanatory variables in the analytical stage. In other words, I would not recommend to exclude Mondays or rainy days, but I would record the date of the interview and whether it was raining or not on that day.

- The collaboration of the health staff on the implementation of the proposed process is essential. It is important, for instance, that on the day of the survey no one without a numbered card is attended. Field-testing should be conducted to define the best ways of ensuring that collaboration.
- The present should to be given to all clients, not just to those that will be interviewed. Therefore it cannot have a high intrinsic value (to facilitate survey management it should ideally have no market value whatsoever), but it should be attractive enough to be an incentive for the clients not to leave the facility before returning their cards to the interviewer. Possible alternatives could be coloured brochures with advise on health practices, calendars, maps of Mozambique, etc.

Selection probabilities and raising factors

The sample of facilities contains 14 explicit strata: the rural area as a whole and each of the 13 urban districts. The probability p_{hi} of selecting district hi in stratum h is 1 for all urban districts, whereas for rural districts it is given by $p_{hi} = r_h n_{hi} / 1,054$, where r_h is the number of rural districts selected (as said before, this will be a number between 13 and 20) and n_{hi} is the number of facilities recorded for district hi in the sample frame.

The probability p_{hij} of selecting facility hij in district hi is given by $p_{hij} = p_{hi} 3 / n'_{hi}$, where n'_{hi} is the number of facilities actually found in the district at the time of the survey.⁴

The raising factor (or weight) w_{hij} that needs to be applied to the data from facility hij in order obtain unbiased national estimates from the sample is the inverse of its selection probability p_{hij} . If, for all rural districts, the actual number n'_{hi} of facilities found at the time of the survey were equal to the number n_{hi} of facilities recorded in the sample frame, the rural sample would be self-weighted, with a constant raising factor equal to $1,054 / (3 r_h) = 351.3 / r_h$. In practice all the numbers will probably not be equal, but they will often be very similar, meaning that the rural sample will not be exactly self-weighted, but quite approximately so.

For the staff sample, the probability p_{hijk} of selecting staff member $hijk$ in facility hij is given by $p_{hijk} = p_{hij} 3 / m_{hij}$, where m_{hij} is the total number of staff in the facility. The corresponding raising factor is $w_{hijk} = 1 / p_{hijk}$.

For the client sample, the probability π_{hijl} of selecting client $hijl$ in facility hij is given by $\pi_{hijl} = p_{hij} c_{hij} / C_{hij}$, where c_{hij} is the number of interviews to clients conducted in the facility on the day of the survey and C_{hij} is the total number of clients who visited the facility that day. The corresponding raising factor is $\omega_{hijl} = 1 / \pi_{hijl}$.

Linkage with the QUIBB survey

The ETSDS may want to link some of its conclusions on to the delivery of health services with the demand for health services observed by the 14,500-household Core Welfare Indicators Survey (QUIBB) that is being conducted concurrently by the Instituto Na-

⁴ If less than three facilities are found in a district at the time of the survey, they will of course all have to be visited, and in that case $p_{hij} = p_{hi}$.

cional de Estatística (INE). The large sample size of the QUIBB survey practically guarantees that such analyses will be possible at all levels that are meaningful.

In the 13 districts that have been classified as “urban” for the purposes of the ETSDS, the QUIBB will indeed observe 3,600 households. In fact, the QUIBB will visit at least 100 households in most of them, an as many as 1,375 households in Maputo alone.

Observing the health-related behaviours of the households in the rural districts visited by the ETSDS will certainly also be possible from the QUIBB. Although the sample of rural districts that will be visited by the ETSDS has not yet been designated, the *pps* procedure that is proposed for their selection virtually ensures that most of them will be largely populated districts where the QUIBB will also observe a fairly large number of households.⁵

Annex I. The sample frame

This annex describes the contents of workbook “Frame.xls”, which should be considered as an integral part of this report.

At the time of this mission two lists of health facilities in Mozambique were available at the Ministry of Health. The first list – reproduced in spreadsheet “Rede A” – contains information on classification, basic infrastructure and staff (the later broken up into four levels) for 1,141 facilities in all 11 provinces. The second list – reproduced in spreadsheet “Rede B” contains more detailed information for 951 facilities in 10 of the 11 provinces (all provinces except Manica). The two lists, however, are not consistent, neither in the identification of the facilities that are listed nor in the attributes reported for the facilities that are present in both lists. Although the facilities are not identified in the lists with standard codes (just by their names), an effort was made to develop a common list, which is reproduced in spreadsheet “Rede A+B”.

As explained in the body of this report, the list in “Rede A+B” is possibly the most complete roster of facilities available at present, but it is not reliable enough to directly select from it the facilities to be visited by the survey. It can be used, however, to estimate the number of primary health care facilities in each district, in order to conduct the first sampling stage (the selection of districts) in the rural areas (the problem does not present itself in urban areas since all urban districts will be visited). Spreadsheet “Distritos” contains the information and formulae required for this process. This spreadsheet contains 142 rows (one for each district), and the first columns contain summary information on demographics (from the census) and on health infrastructure (from spreadsheet “Rede A+B”).

⁵ Simulations indicate that about 100 QUIBB households will be found in each ETSDS district on the average.

The screenshot shows a Microsoft Excel spreadsheet with the following data:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
	Prov Code	Provincia	Dist Code	Distrito	Region	Área	Homens	Mulheres	Total Pop	Health Facilities	Health Staff	FLCF Count	FLCF Staff	
2	01	Niassa	0102	Cuamba	N	R	81,555	84,450	166,005	12	71	11	9	
3	01	Niassa	0103	Lago	N	R	33,627	34,821	68,448	11	21	11	21	
4	01	Niassa	0104	Lichinga Distrito	N	R	35,106	36,353	71,459	13	16	13	16	

Columns “Homens”, “Mulheres” and “Total Pop” contain the population of the district by gender, columns “Health Facilities” and “Health Staff” contain the total number of health facilities and health staff (in all levels) that are recorded in spreadsheet “Rede A+B”, and columns “FLCF Count” and “FLCF Staff” contain the total number and the staff in primary health care facilities.

Columns “O” to “S” contain the formulae required to select a sample of rural districts with probability proportional to the number of primary health facilities. Only two figures need to be plugged into the spreadsheet to these effects: The number of desired districts (on cell “P1”) and a real random number between 0 and 1 (on cell “Q1”). The selected districts are automatically flagged with a number “1” on column “R”.

The screenshot shows a Microsoft Excel spreadsheet with the following data:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
	Prov Code	Provincia	Dist Code	Distrito	Region	Área	Homens	Mulheres	Total Pop	Health Facilities	Health Staff	FLCF Count	FLCF Staff						
1															1,054	13	0.31		P 1
2	01	Niassa	0102	Cuamba	N	R	81,555	84,450	166,005	12	71	11	9	11	0.14	0.45			
3	01	Niassa	0103	Lago	N	R	33,627	34,821	68,448	11	21	11	21	11	0.14	0.59			
4	01	Niassa	0104	Lichinga Distrito	N	R	35,106	36,353	71,459	13	16	13	16	13	0.16	0.75			
5	01	Niassa	0105	Majune	N	R	10,916	11,303	22,219	6	17	6	17	6	0.07	0.82			
6	01	Niassa	0106	Mandimba	N	R	49,098	50,841	99,939	8	25	8	25	8	0.10	0.92			
7	01	Niassa	0107	Mamupa	N	R	22,881	23,694	46,575	4	16	4	16	4	0.05	0.97			
8	01	Niassa	0108	Maúã	N	R	17,374	17,991	35,365	5	13	5	13	5	0.06	1.03	1	0.06167	
9	01	Niassa	0109	Mavago	N	R	5,560	5,757	11,317	3	7	3	7	3	0.04	1.07			
10	01	Niassa	0110	Mecanhelas	N	R	42,103	43,598	85,701	9	27	9	27	9	0.11	1.18			
11	01	Niassa	0111	Mecula	N	R	6,778	7,018	13,796	13	10	13	10	13	0.16	1.34			
12	01	Niassa	0112	Metarica	N	R	12,739	13,192	25,931	6	6	6	6	6	0.07	1.41			

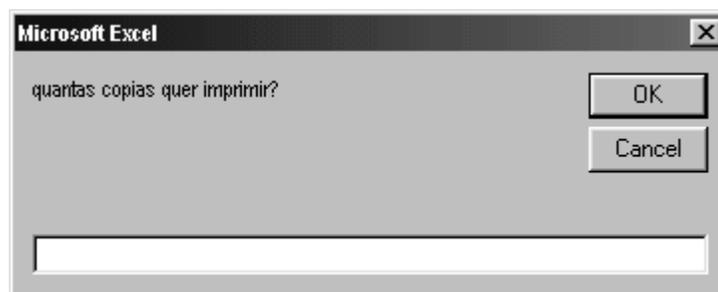
The detailed contents of columns “O” to “S” is as follows:

- Column “O” is the measure of district size that will be used for the *pps* selection. In this case, it is just a straight copy of column “L” (the number of primary health facilities in the district), but it could be replaced by alternative measures such as the district population or the number health staff. Cell “O1” is the total size of all rural districts.
- Cell “P2” contains the formula “ $O2/O\$1 * P\1 ”, and it is copied all the way down column “P”. It represents the “scaled size” of the district: its size scaled by a factor such that the total size of all rural districts equals the number of districts to select (the contents of cell “P1”).
- Cell “Q2” contains the formula “ $Q1+P2$ ”, and it is copied all the way down column “Q”. It represents the “cumulated scaled size” of the districts, shifted by the random number in cell “Q1”.
- Cell “R2” contains the formula “ $INT(Q2) - INT(Q1)$ ”, and it is copied all the way down column “R”. It will flag the districts selected for the sample, by yielding a “1” whenever the integer part of column “Q” changes to the next integer (or zero otherwise).
- Cell “S2” contains the formula “ $P2 * R2$ ”, and it is copied all the way down column “S”. It represents the selection probability of each rural district in the sample (referred to as p_{hi} in the body of the report).

Annex II. Forms for sample selection

This annex gives drafts of the three forms required to support the sampling procedures that need to be implemented as a part of the survey field operations. All the forms are contained in an Excel workbook called “Formas.xls” that is attached to the original of this report. Each form is in a separate spreadsheet:

All three forms implement some kind of randomisation. Since the forms will be used in relatively few occasions (the selection of primary health facilities will be done at most in 30 districts and the selection of staff and clients in at most 100 facilities), it seems better to centrally print as many *different* forms as needed (rather than reproducing a number of identical forms and ask the survey staff to use tables of random numbers in the field). The workbook contains a macro that facilitates this process: each time a spreadsheet is activated, the following dialog box appears:



Entering the number of required copies in the space provided and clicking on the **OK** button will then automatically print the required number of different copies. If no copies are needed at the moment, click the **Cancel** button or press the **Esc** key.

The form reproduced in the next page (contained in spreadsheet “Unidades”) can be used to prepare the list of all primary health facilities in each district and select from that list a *simple random sample* of three facilities to be visited. The field staff should be instructed to complete the list of units in some clearly specified order (for instance, in order of increasing distance from the district centre), and then use the list of random numbers in the bottom right to select the three facilities to be visited, following the explicit instructions in Portuguese included in the form itself.

Inquérito sobre Serviços Públicos - Sector Saúde, Moçambique

Lista das unidades sanitárias do distrito e selecção das unidades a serem visitadas

Nome do Distrito:	Provincia	Distrito	Responsavel da lista e à selecção:	Dia	Mes	Ano

Número Unidade	Nome da Unidade	A
01		
02		
03		
04		
05		
06		
07		
08		
09		
10		
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		
21		
22		
23		
24		
25		
26		
27		
28		
29		
30		

Número Unidade	Nome da Unidade	A
31		
32		
33		
34		
35		
36		
37		
38		
39		
40		
41		
42		
43		
44		
45		
46		
47		
48		
49		
50		

Usar os números em baixo para seleccionar as unidades sanitárias a serem visitadas no distrito. Ler cada fila da esquerda à direita, começando com a fila superior, até encontrar os números de tres unidades. Não fazer caso dos números maiores que a quantidade total de unidades. Circular erros números e sinalar as unidades com uma cruz na colona A.

30	31	37	46	24	45	10	47	42	26
13	01	34	23	06	25	16	21	09	27
33	22	40	48	44	02	14	32	04	07
11	12	19	05	29	03	17	38	41	08
15	49	50	18	28	20	39	43	35	36

The form reproduced in the next page (contained in spreadsheet “Pessoal”) can be used to prepare the list of all staff in each facility and select from that list a *simple random sample* of three staff members to be interviewed. The field staff should be instructed to write the list of staff members in a clearly specified way (for instance, according to the professional levels defined by the Ministry of Health) and then use the random numbers to choose the three staff members.

Inquérito sobre Serviços Públicos - Sector Saúde, Moçambique

Lista do pessoal da unidade e seleção das pessoas a serem entrevistadas

Nome da Unidade:	Provincia	Distrito	Número

Responsável da lista e à seleção:	Dia	Mes	Ano

Número Pessoa	Nome da pessoa	Posição	A
01			
02			
03			
04			
05			
06			
07			
08			
09			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23			
24			
25			
26			
27			
28			
29			
30			

Número Pessoa	Nome da pessoa	Posição	A
31			
32			
33			
34			
35			
36			
37			
38			
39			
40			
41			
42			
43			
44			
45			
46			
47			
48			
49			
50			

Usar os números em baixo para seleccionar as pessoas a serem entrevistadas na unidade. Ler cada fila da esquerda à direita, começando com a fila superior, até achar os números de tres pessoas. Não fazer caso dos números maiores que a quantidade total de pessoas na unidade. Circular esses números e sinalar as pessoas com uma cruz na colona A.

30	31	37	46	24	45	10	47	42	26
13	01	34	23	06	25	16	21	09	27
33	22	40	48	44	02	14	32	04	07
11	12	19	05	29	03	17	38	41	08
15	49	50	18	28	20	39	43	35	36

Inquérito sobre Serviços Públicos - Sector Saúde, Moçambique

Lista dos utentes no dia do inquerito e seleção dos utentes a serem entrevistados

Nome da Unidade:	Provincia	Distrito	Número	Responsavel da lista e à seleção:	Dia	Mes	Ano

Quantidade de utentes esperados:	/ 6 =	(omitir decimais)	x	0.965	=	(omitir decimais)	+ 1 =	Número do primeiro utente a serem entrevistado:

Número Utente	Sinalar com una cruz quando		
	Recebeu o presente	Tem que ser entrevistado	Foi entrevistado
01			
02			
03			
04			
05			
06			
07			
08			
09			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23			
24			
25			
26			
27			
28			
29			
30			

Número Utente	Sinalar com una cruz quando		
	Recebeu o presente	Tem que ser entrevistado	Foi entrevistado
31			
32			
33			
34			
35			
36			
37			
38			
39			
40			
41			
42			
43			
44			
45			
46			
47			
48			
49			
50			
51			
52			
53			
54			
55			
56			
57			
58			
59			
60			

Número Utente	Sinalar com una cruz quando		
	Recebeu o presente	Tem que ser entrevistado	Foi entrevistado
61			
62			
63			
64			
65			
66			
67			
68			
69			
70			
71			
72			
73			
74			
75			
76			
77			
78			
79			
80			
81			
82			
83			
84			
85			
86			
87			
88			
89			
90			

Número Utente	Sinalar com una cruz quando			Número Utente	Sinalar com una cruz quando			Número Utente	Sinalar com una cruz quando		
	Recebeu o presente	Tem que ser entrevistado	Foi entrevistado		Recebeu o presente	Tem que ser entrevistado	Foi entrevistado		Recebeu o presente	Tem que ser entrevistado	Foi entrevistado
91				131				171			
92				132				172			
93				133				173			
94				134				174			
95				135				175			
96				136				176			
97				137				177			
98				138				178			
99				139				179			
100				140				180			
101				141				181			
102				142				182			
103				143				183			
104				144				184			
105				145				185			
106				146				186			
107				147				187			
108				148				188			
109				149				189			
110				150				190			
111				151				191			
112				152				192			
113				153				193			
114				154				194			
115				155				195			
116				156				196			
117				157				197			
118				158				198			
119				159				199			
120				160				200			
121				161				201			
122				162				202			
123				163				203			
124				164				204			
125				165				205			
126				166				206			
127				167				207			
128				168				208			
129				169				209			
130				170				210			

M:\PS Document

Library\PSwebsite\Mozambique\Documentation\mozambique.samplingnote.JM.sept4.doc

September 4, 2002 5:44 PM