48930
V

UNITED NATIONS
ECONOMIC AND SOCIAL COUNCIL

Distr.
LIMITED
E/CN.14/SM/37
4 May 1981
Original: ENGLISH

ECONOMIC COMMISSION FOR AFRICA

Working Group on Organization, Content and Methodology of Household Surveys

Addis Ababa, 29 June - 3 July 1981

SURVEY DESIGNS OF FOUR COUNTRIES UNDER AFRICAN HOUSEHOLD SURVEY CAPABILITY PROGRAMME

C O N T E N T S

M81-1511

SURVEY DESIGNS OF FOUR COUNTRIES UNDER AFRICAN HOUSEHOLD SURVEY CAPABILITY PROGRAMME

I. INTRODUCTION

1. The National Household Survey Capability Programme (NHSCP) in which the African Household Survey Capability Programme (AHSCP) is a regional component aims to help the developing countries to obtain vital information which they need for their national development policies and programmes through household surveys in conjunction with data from other sources. It is designed to build up durable national facilities for carrying out household surveys and to achieve self-reliance in national statistics.

2. The programme envisages a continuous and co-ordinated series of national household surveys to generate integrated data overtime on a wide range of subjects. Integration of data can be secured between subjects spread over different rounds through common and comparable concepts and definitions, and methods of data collection and analysis. Linkages can be established through a common set of core items to be repeated in successive rounds. The objectives of integration of data are not only to assess the current levels of a number of socio-economic phenomena and the interrelationships among them, but also to study the trends and changes in relationships over time.

3. A question arises in relation to the level of integration of data over different survey rounds as to whether it is possible to link data at the household level or at a higher level. This question can be considered in the light of a country's experience, resources available and needs, and the sample design plays a crucial role in this regard. Hence in formulating a sample design for a country in the AHSCP framework, the level of integration of data is one important point to be reckoned with in addition to the other factors such as background and features of the country, sampling and non-sampling errors, cost, domains of study and available resources.

4. This paper deals with four African countries which are participating in AHSCP and have developed survey designs to generate integrated data. In the first part the countries' background, features, survey experience and programme and sample designs used or to be used are described. The advantages and disadvantages of these designs will be briefly pointed out. Finally a comparative study of survey programmes, sample designs and capabilities will be made.

II. BOTSWANA

5. The Republic of Botswana covers an area of 582,000 square kilometers with the Kalahari desert occupying a large part of the central and southern area of the country. The estimated population (1980) is about 0.8 million distributed unevenly with 80 per cent of the population living in the eastern region. The urban and peri-urban population is estimated to constitute about 15 per cent of the total. The country is divided into nine administrative districts, three towns and one capital.

6. The Central Statistics Office (CSO) carried out the population census in 1971 and has conducted a number of ad hoc household surveys since 1968 and annual livestock and crop survey jointly with the Ministry of Agriculture. The population census to be undertaken in 1981 will provide a reliable frame for the sample surveys. However, the CSO had no permanent field staff and the census and surveys were undertaken using a few supervisors of the CSO and temporary staff.

7. The survey programme for Botswana under the AHSCP, from 1981 to 1985 will cover:

(a) Annual survey on agriculture and livestock.
(b) Income, consumption and expenditure incorporating nutrition.
(c) Labour force and migration.
(d) Disposal of household production.
(e) Fertility, mortality and family planning, including questions on social perspectives (Demographic Survey).

Certain core items reflecting the most important demographic, social and economic characteristics of the people of Botswana will be repeated from one survey to another to provide linkage between various surveys.

8. A stratified two-stage sampling based on the experience of the National Migration Survey carried out in Botswana will be applied in both rural and urban areas as follows.

9. In the rural survey outside the freehold farm areas, a primary sampling unit will be a village catchment area consisting of 1 to 17 census enumeration areas and in the freehold farm areas it will be a freehold farm. A secondary sampling unit is a household in all rural areas. The stratification parameters would vary according to the theme of the different modules or subjects in the programme.

Theme		Stratification Parameters
Demographic/Labour force	–	Population density
Agricultural Production	–	Agro-climatic zones
	–	Cropping patterns
	–	Herding practices
Income/Expenditure	–	Income
	–	Herd size
Health/Nutrition	–	Consumption patterns

10. A sample of 1500 households will be drawn from the rural survey by selecting the primary sampling units with probability proportional to size of population. The details on selection of units from each stratum and in each stage have not been specified.

11. In the urban survey, it is proposed to use a cluster of housing units (block/plot) as a primary sampling unit and a dwelling unit/house as a secondary sampling unit. The PSUs will be stratified on the basis of type of housing and so the following strata are identified in the design.

High cost
High medium cost
Medium cost
Low cost
Site and service
Traditional
Peri-urban: up-graded
squatter
Domestic servants quarters

12. An optimum size of a cluster will be determined through a statistical experiment to make the design efficient. A sample of 400 households will be selected in the second stage assuming that there is a one-to-one relationship between a dwelling unit and a household.

13. Combination of rural and urban samples produces a sample of 2000 households which will be enumerated approximately every four months giving rise to three survey rounds in a year, as dictated by the requirements of the annual agricultural production survey.

14. It is proposed to use a replicated systematic design by which 2000 households in each round are divided into four replicates of 500 households - 400 rural and 100 urban households. By this scheme, the sample will be rotated within a year or every year by replacing two replicates in order to minimise the non-sampling errors due to respondents.

15. Advantages:

(a) The 1981 Population Census will provide a reliable frame and weights for the selection of the sample with unequal probability.
(b) The use of a replicated systematic design with a rotating scheme can reduce non-sampling errors due to respondents.
(c) The statistical exercise proposed to be conducted to determine the size of cluster in the urban survey can produce the interesting intra-class correlations of different variables.

16. Disadvantages:

(a) Since a master sample has not been proposed, the frame of all primary units needs to be updated whenever a new sample is selected.
(b) It is not clear why the village catchment area in the rural survey and a cluster of housing units (block/plot) in the urban survey rather than a census enumeration area are proposed to be used as primary sampling units. Similarly there may not be a one-to-one relationship between dwellings/houses and households in the urban surveys.
(c) If the units are stratified by different parameters for different modules, presumably in different rounds, integration of data will become difficult at least at the lower level.
(d) A sample of 2000 households in each year might be sufficient to produce reliable estimates at the national level and possibly provincial level for some variables but scarcely for the others like demographic variables.

III. CAMEROON

17. The United Republic of Cameroon has a surface area of 475,442 square kilometers and an estimated population of 3.4 million. It is administratively divided into 40 districts and 147 sub-districts (arrondissements) but ecologically divided into six zones as Cocoa, Tobacco, Coffee, Cotton, Livestock and Unclassified zones. There are two large cities Yaounde and Douala.

18. The Department of Statistics and National Accounts (DSCN) has carried out many ad hoc surveys including National Fertility Survey under the WFS Programme and a population and housing census. The last census of 1976 would provide a frame including the maps of enumeration areas and list of population and households in those EAs.

19. The continuous and integrated surveys to be carried out by the DSCN under the AHSCP are programmed as follows:

1981-82	–	Survey of income, consumption and expenditure with special modules on food consumption and nutrition.
1982-83	–	Survey on employment, including modules on the participation of women in development and on activities in the formal sector.
1983-84	–	(a) Health and environmental survey with module on the handicapped. (b) Demographic and housing survey.
1984-85	–	Survey of social perspectives with special module on rural development.

20. Although the programme was scheduled to start in 1979, it could only be initiated in 1981 due to some reasons, the major one of which was the unavailability of external assistance. The Government of Cameroon is financing the first survey on income, consumption on its own with technical backstopping of the ECA. To support the remaining surveys, negotiations have started with potential donors.

21. The Government needs data separately from the rural and urban areas of ecological zones and two cities. Two ecological zones being small are combined with the neighbouring zones which have some similar characteristics with them so that there are ten domains of study, namely:-

1-2 Urban and rural areas of cocoa-tobacco zone
3-4 Urban and rural areas of coffee zone
5-6 Urban and rural areas of cotton-livestock zone
7-8 Urban and rural areas of unclassified zone
9 Yaounde
10 Douala

22. Of 147 sub-districts, 7 sub-districts which are sparsely populated and not easily accessible (with less than 10 per cent of the total population) are dropped out of the frame. The sample designs are formulated taking into account:

(i) cost, particularly cost of transportation
(ii) control of non-sampling errors
(iii) obtaining fairly reliable estimates for each domain
(iv) simplified method of data processing and tabulation

23. In each ecological zone (or combination of two zones), a three-stage stratified sampling is applied using a sub-district (arrondissement), a segment of about 120 households and a household as units in the respective stages. Since the population of sub-districts are unequal, a sample of them is selected by systematic sampling with probability proportional to size. A selected sub-district is then stratified into urban and rural areas and a number of segments is selected from each stratum by circular systematic sampling. Finally a fixed fraction of households is selected by circular systematic sampling using two starting points to get two independent replicated sub-samples in each segment.

24. The number of segments to be chosen is fixed such that the design is self-weighting and at the same time the sample is sufficiently large to obtain reliable estimates. Since towns and rural areas in the sample belong to the same sub-districts, cost and non-sampling errors can be controlled by allocating the vehicles and field staff more flexibly and efficiently.

25. Since an average size of a segment is around 120 households out of which 24 households are aimed to be selected (i.e., 4 households per day to be enumerated by an interviewer), the sampling fraction for each sub-sample is 1/10 and total sampling fraction is 1/5. The replicated sub-samples will enable application of sampling with partial replacement in the following surveys.

26. In Yaounde and Douala, a stratified two-stage sampling is applied by taking a sub-district as a stratum, a segment (of about 60 households) and a household as primary and secondary sampling units. From each stratum, two replicated samples are selected by circular systematic sampling. And two replicated sub-samples of households are again selected from each segment in the same manner, but at constant sampling fraction to make the design self-weighting.

27. In these two cities, about 12 households are aimed to be selected from each segment, so the sampling fraction for each sub-sample is 1/10. An interviewer will be responsible to survey two segments per week (i.e., 4 households per day).

23. The whole sample will be consisted of 33 sub-districts (out of 140) and 303 segments and 6048 households to be enumerated by the field team of 22 supervisors and 84 interviewers. In the income, consumption and expenditure survey, a rotating sample each covering one third of the sample with monthly replacement and quarterly revisit will be used so that 2016 households will be surveyed every month (i.e., 24 households by each interviewer in each month with weekly visit).

29. A master sample will be constructed based on the selected sample which will be treated as a sub-sample. The samples for the following surveys will also be drawn as sub-samples from the master sample. The expected advantages of having master sample for Cameroon, as in the other countries, are reduction in cost of up-dating the frame and integration of data from different rounds. And by replacing the sample partially at the lower stage, non-sampling error due to respondents can also be controlled.

30. Advantages:

(a) Design is based on a frame of census maps.
(b) Cost of transportation will be reduced.
(c) Non-sampling error will be controlled by making field staff mobile in each selected sub-district.
(d) Fairly reliable estimates for domains are expected.
(e) Data processing and tabulation are simplified by having self-weighting designs.
(f) A master sample will be constructed. Sampling with partial replacement will be applied in the following surveys to control some non-sampling errors due to respondents.

31. Disadvantages:

(a) Since the frame is based on the 1976 population census, lists of population and households in the EAs are out-of-date.
(b) By making design self-weighting, allocations of segments in the sub-districts and strata are not proportional.
(c) Since the samples for the rural and urban areas of ecological zones are selected with probability proportional to the population of the whole sub-district, precision of estimates for each area might not be as high as those estimates based on samples selected separately.
(d) Work-load for interviewers will not be equally distributed although it will not vary highly on the average.

IV. ETHIOPIA

32. Ethiopia is a large country in East Africa with a total land area of 1.22 million square kilometers and an estimated population of around 30 million. Most of the people live in the highland areas and about 10 per cent of the population are nomadic living mainly in the southern and eastern areas. Only 12 per cent of the population are residing in urban areas.

33. The country is divided into 14 regions and 102 sub-regions (Awrajas) and 583 districts (Woredas). The lowest administrative unit is the Farmers' Association (FA) in the rural area and the Urban Dwellers' Association (UDA) in the urban area. Over 23,000 FA's and 1133 UDA's have been formed with the exception of nomadic areas. One FA contains about 250 households and covers up to 300 hectares and one UDA consists of about 500 households.

34. Ethiopia has never undertaken a census in the modern sense but is preparing to launch the first one in 1983 using proper cartographic maps which will furnish the surveys with a reliable frame.

35. The Central Statistical Office (CSO) of Ethiopia conducted two rounds of National Sample Surveys (1964-67) and (1968-71) and many other surveys covering many subjects, most of which were of ad hoc nature. The CSO in 1980 prepared a National Integrated Household Programme (NIHSP) consisting of the following surveys, population census and data collection:

(1)	Annual crop production and livestock survey	
(2)	Socio-demographic survey (including physical disability)	- 1981 - rural area - 1982 - urban area
(3)	Income, consumption and expenditure survey (including summary labour force survey)	- 1981-82
(4)	Labour force survey	- 1982-83
(5)	Nutrition survey	- 1982-83
(6)	Population census	- 1983
(7)	Post Enumeration Survey	- 1983
(8)	Urban household small scale enterprise survey	- 1983
(9)	Socio-demographic survey (household amenities and social services)	- 1984
(10)	Large scale agricultural survey	- 1984-85
(11)	Socio-demographic survey (literacy and education)	- 1985
(12)	Income, consumption and expenditure survey	- 1985-86
(13)	Nutrition survey	- 1985-86
(14)	Socio-demographic survey (including position of women and children)	- 1986
(15)	Price data	
(16)	Community level data	
(17)	Quantitative data on crop and food situation	

36. In the first few years, it is aimed to obtain reliable estimates at national level with urban and rural disaggregation and meaningful estimates at regional and possibly lower level. The programme of survey will cover the settled population only in the first few rounds. Price data will be collected by the enumerators on the market days and community level data and quatitative data on crop and food situation will be gathered by the supervisors.

37. The frame of Farmers' Associations has been prepared and up-dated by the Central Statistical Office recently. Although there are no proper maps, the average area of a FA is known to be about 800 hectares and the number of FA members which corresponds to the number of holdings are available for each FA. Hence a FA and a member or holding are taken as primary and secondary sampling units.

38. A stratified two-stage sampling is applied in the rural survey by taking a sub-region (Awraja) as a stratum. A total sample of 500 FA's are first allocated to regions and then to strata in proportion to the number of FA members. Then FA's are selected with probability proportional to size (i.e., number of members or holdings) with replacement in the first stage.

39. In agricultural survey, 25 FA members are selected from each FA with equal probability and data are collected by the interview method. Some of the fields under different crops are selected at random for area measurement and two plots at random out of each field are selected again for crop cutting.

40. In the demographic survey, more households will be selected from each FA to raise the sample size to 100. The additional households will be selected with PPS of number of FA members in the households. Later on, some non-farming households will be added to the sample. In the other surveys, the same households selected for agricultural survey and if necessary, with additional households will be enumerated.

41. The first population census and a large scale agricultural survey will be conducted in 1983 and 1984-85, for which large field organizations and other resources will be mobilised.

42. In the urban areas, each city or a large regional capital will be taken as a stratum and the other towns will constitute another stratum. A two-stage sampling will be applied in each city or a capital by selecting 18 UDAs with probability proportional to the number of UDA members in the first stage and households with equal probability in the second stage. In the stratum of other towns, a three-stage sampling will be applied to select the towns with PPS in the first stage and 54 UDAs with PPS and households with equal probability in the following stages.

43. The households selected from the rural and urban areas will be retained until the census cartographic maps and more information are available in 1983. Then the frame will be up-dated and a new and bigger sample will be selected covering a larger area using a new design and larger field staff.

44. The field work is organized on regional basis headed by the Deputy Administrator of the Regional Planning Supreme Council under whom there are Chief Supervisor, Supervisors and enumerators. An enumerator is posted in each FA or one or two UDAs throughout one survey round.

45. Advantages:

(a) The sample design in the rural area is dictated by the need to obtain annual crop production and livestock data.
(b) As the sample consists of 500 and 100 primary sampling units in the rural and urban surveys respectively reliable national estimates can be expected.
(c) In the rural sample since 34-100 FA's are selected from each region, estimates at that level should be reliable.
(d) Field survey work is properly organized.

46. Disadvantages:

(a) As the programme consists of many surveys and an overlapping population census there can be problems at all stages.
(b) The first frame is not based on census maps and data.
(c) The design can be more efficient if the primary sampling units are selected with PPS without replacement.
(d) Since the households are selected with probability proportional to members of Associations, the design may not be suitable for the socio-demographic and economic surveys. There can be some difficulty in tabulation.
(e) If sample is not rotated, there can be non-sampling errors due to respondents.

V. KENYA

47. The Republic of Kenya covers a land area of about 0.57 million square kilometers of which 0.1 million square kilometers are considered to be arable. Kenya's provisional population as of August 1979 was 15.3 million of which 85 per cent were estimated to be in rural areas.

48. The country is administratively divided into 7 provinces and 40 districts. The North Eastern Province is large in area but sparsely inhabited by nomadic or semi-nomadic people accounting for 5-7 per cent of total population of the country. It has been difficult to cover these people in the surveys conducted in the past.

49. The Central Bureau of Statistics (CBS) of Kenya completed in 1979 their first phase of National Integrated Sample Survey Programme (NISSP) 1975-79, and conducted a population census in August 1979. On the basis of the frame of Enumeration Areas (EAs) of the 1979 census, a sample design for the next round of national surveys called National Sample Survey and Evaluation Programme (NASSEP) is drawn. The NASSEP of 1980-84 will comprise the following main surveys.

Year	Rural	Urban
1980-81	Selection of national sample and post-enumeration survey of Population Census, 1979. Literacy and handicapped surveys.	Selection of national sample and post-enumeration on survey of Population Census, 1979.
1981-82	Budget (income, consumption and expenditure) and Agricultural Production Survey.	Budget survey, literacy survey, handicapped survey.
1982-83	Survey of health, nutrition and social indicators.	Survey of health, nutrition and social indicators.
1983-84	Agricultural Production Survey, Demographic Survey (including fertility, mortality and migration).	Demographic Survey (including fertility, mortality and migration).
1984-85	Employment survey	Employment survey

50. The aim of NASSEP is to provide the necessary data for each district or combination of very sparsely populated districts and for each major urban centre and groups of other urban centres. In addition to the above programme, the CBS is developing a capacity to conduct ad hoc or continuous surveys to evaluate the impact of specific development projects.

51. The rural survey design is a stratified three-stage sampling taking a district or combination of some of them as a stratum. The primary sampling unit is a census EA whose size will be calculated in units of nearest 100 households, denoted by (k). From each of 27 strata 24 PSUs will be selected by systematic sampling with probability proportional to assigned size (k) of units. In the second stage if the selected EA has size k [illegible], then it will be divided into k secondary sampling units and one is selected at random so that probability of selection of a SSU is constant. In the ultimate stage a constant fraction of households from each secondary sampling unit will be selected to make the design self-weighting. At first it was considered to use only one third of the sample of SSUs in the budget survey but it has now been decided to use the whole sample.

52. The sample will be selected by a similar stratified three-stage design in the large urban centres. The other urban areas will be grouped into some strata possibly by province and the details of the design to be applied in those areas have not been specified.

53. The North Eastern Province and two districts in Rift Valley Province, excluded from the national sample are to be dealt with in an ad hoc manner in terms of sample design. Surveys of livestock which may well be an important feature of the programme may make use of grid or strip samples using aerial phtography. Surveys involving interviews with individual respondents will require a listing of the population within a given primary unit just prior to the survey in view of the mobility of the population - where such mobility exists.

54. Kenya plans to construct a master sample and the survey is considered to be designed for the urban and rural areas of the master sample. On the other hand this national sample will evolve into a rotating sample whereby each year a proportion of primary sampling units are replaced by a new selection of primary units - possibly one-quarter each year. The reason for such a replacement is to avoid inherent risk of having the same household in several surveys.

55. Advantages:

(a) The frame of EAs is reliable being based on the maps and results of 1979 census.
(b) The design has been formulated to be self-weighting at stratum level to simplify data processing and tabulation. A precaution has been taken to minimise errors in the "measure of size" of secondary units as encountered in the previous integrated surveys.
(c) A master sample will be constructed.
(d) A rotating sample will be employed in each round or year to reduce respondents' error.

56. Disadvantages:

(a) It is possible to reduce the sample design by one stage by selecting the clusters of households in the first stage to make the design more efficient.
(b) If the primary sampling units are partially replaced every year or every round, it will be expensive.
(c) If the size of districts or strata varies, the units would rather be allocated either by proportional allocation or optimum allocation to make the design more efficient and less costly.
(d) Since a domain of study is a district, there can be non-sampling errors if the field work is not well organized.

VI. COMPARATIVE STUDY ON SURVEY PROGRAMME, DESIGN AND CAPABILITY OF FOUR COUNTRIES

57. Table 1 gives area and population of Botswana, Cameroon, Ethiopia and Kenya along with their survey experience, programme design, sample units and sizes, integration of survey results, field staff and technical staff, training programme, expected cost, data processing, cartographic and printing capability and major constraints.

58. In comparing these countries, Botswana, Cameroon and Kenya are of about the same size and Ethiopia is the largest in area and population and Botswana is the smallest in terms of population. As for experience in censuses and surveys, Kenya is at the forefront having carried out censuses and a National Integrated Survey Programme (1975-79).

59. The survey programme of Ethiopia is the most ambitious one having many surveys to be undertaken every year and the first population census to be carried out during that period. Kenya also has a programme involving many surveys and Botswana plans to cover annual core variables on many subjects. Cameroon is the only country with a moderate programme.

60. Regarding level of estimation, Botswana and Ethiopia aim to have reliable national and provincial estimates in the first few years and Cameroon aims to obtain reliable estimates at the level of zone and city. On the other hand Kenya plans to obtain estimates at district and city levels and other urban areas.

61. The frames for the surveys are based on the census maps and data in all countries except in Ethiopia. In that country the census of 1983 will provide a new frame and then the sample design will be revised accordingly.

62. Caemroon and Kenya will establish a master sample whereby only the units in that sample need to be updated for every survey round. But in Kenya if the sample of primary units is to be replaced partially every year or every round, a large master sample will be required to be updated and then it will be a costly operation.

63. The sample designs of the four countries are different from each other depending on their priorities, available resources and experience. Since these designs have been described above, only the salient points will be discussed.

64. Botswana tries to integrate the agriculture and livestock survey into the socio-economic survey programme. It plans to replace two replicated samples every year or round. This arrangement will enable it to reduce some non-sampling errors due to respondents but will be expensive if the new EAs are selected every year. And stratification based on varying parameters for different modules and various types of housing can cause problems in selection of sample and integration of surveys.

65. In Cameroon, two replicated samples of households are selected from each segment so that each of them can be replaced after each survey. The other aim of the design is to reduce cost and non-sampling error by selecting segments of urban and rural areas from the same sub-district. This arrangement can increase sampling error so it becomes a matter of trade-off between cost and non-sampling error against sampling error.

Table 1

Description of countries, their survey programme, design and capability

No.	Particulars	Botswana	Cameroon	Ethiopia	Kenya
1	Area (m.sq.km.)	0.53	0.47	1.22	0.
2	Population (m.)	0.8 (1980)	7.63 (1976)	30.0 (est.)	15.3 (1979)
3	Latest census	1981	1976	1983	1979
4	Survey experience	Ad hoc	Ad hoc	Ad hoc	National Integrated Programme (1975-79)
5	Period of survey programme	5 years	5 years	5-6 years	5 years
6	Survey programme***	(a),(b),(c),(d),(e),(f),(g),(h),(n),(o)	(b),(c),(d),(e),(f),(h),(j),(l),(o)	(a),(b),(c),(d),(e),(g),(h),(i),(j),(k),(l),(m),(n),(o)	(a),(b),(c),(d),(e),(f),(h),(i),(j),(k),(l),(o),(p)
7	Frame	1981 census	1976 census	List of units to be revised when maps are available.	1979 census
8	Master sample	No	Yes	No	Yes
9	Level of estimation	National, Rural, Urban	Rural and Urban areas of ecological zones and two cities	National, Regional and Urban	District, cities and urban areas
10	Survey design	Stratified two stage sampling with PPS (Replacement of 2 replicated samples out of 4 every year or round).	Stratified three stage with PPS in each zone; Stratified two stage in each city (Sampling with partial replacement of households or segments after each round).	Stratified two stage with PPS with replacement in rural area; Stratified two stage with PPS in cities and capitals; Three stage with PPS in other urban areas (Fixed sample for the first two years to be revised afterwards).	Stratified three stage with PPS in each district and city; Stratified four stage with PPS in the other urban areas (Rotating sample with partial replacement of PSUs).

No.	Particulars	Botswana	Cameroon	Ethiopia	Kenya
11	Self-weighting	Uncertain	Yes	Uncertain	Yes
12	Sampling units	Rural PSU - EA/VCA* SSU - Household	Zone (Rural)+ Urban) PSU - sub-district SSU - segment USU*- household	Rural PSU - Farmers' Association SSU - member of FA or household	District and Cities PSU - EA SSU - Cluster of 100 households USU - household
		Urban PSU - Block (plot) SSU - dwelling unit	Two cities PSU - segment SSU - household	Cities PSU - Urban Dwellers' Association SSU - household	Other Urban PSU - town SSU - EA TSU - Cluster USU - household
		*VCA - Village Catchment Area consisting of some EAs	*USU = Ultimate Stage Sampling Unit		
				Other Urban PSU - town SSU - UDA USU - household	
13	Sample size	Rural 1600 households	Rural 3456 households	Rural 12500 FA members	Rural and Urban 10,000 households
		Urban 400 households	Other Urban Area 1224 households Two Cities 1326 households	Urban 5912 households	
14	Integration of Survey results	Possible	Possible	Possible	Possible
15	Field staff	7 supervisors 21 interviewers (additional) 4 supervisors 12 interviewers for income and expenditure survey)	22 supervisors 34 interviewers 12 editors for income and expenditure survey)	120 supervisors 600 interviewers (to be increased up to 140 supervisors 720 interviewers)	30 supervisors 300 interviewers 145 casual staff

No.	Particulars	Botswana	Cameroon	Ethiopia	Kenya
16	Technical staff				
	External	6	0	8	3**
	National	6	10	82	30**
	Total	12	10	90	33
17	Training Programme	Yes	Yes	Yes	Yes
18	Expected cost ($million)				
	External	1.05 (52%)	1.77 (44%)	2.35 (20%)	2.15 (18%)
	Government	0.98	2.17	11.39	9.91
	Total	2.03	3.94	14.24	12.06
19	Average cost per annum ($million)	0.41	0.79	2.76	2.41
20	Computer in the Statistics Office	No computer Have terminals linked with ICL 2956	IBM 32 connected with IBM 370 at Government Data Processing Centre	NCR 8455	Wang 2200-MPV Accessible to IBM 370/135 DOS/VS
21	System analysts/ programmers.	None	2	Some	Some
22	Cartographers	1	5	10	Some
23	Printer	No	No	Yes	Yes
24	Major constraints	Field staff, transportation and data processing	Transportation and data processing	Transportation and data processing	Transportation and data processing

** One system analyst (for 42 man-months)
One or two consultants (for 12 man-months)
15 national staff (full-time)
15 national subjects specialists (part-time)
(Assumption)

*** (a) Annual agriculture and livestock survey
(b) Income, consumption and expenditure survey
(c) Labour force (and migration) survey
(d) Demographic survey
(e) Social perspectives
(f) Health (and environmental) survey
(g) Household enterprise survey
(h) Nutrition survey
(i) Literacy (education) survey
(j) Handicapped or disability survey
(k) Market survey
(l) Agriculture census or large scale agriculture survey
(m) Population census
(n) Post enumeration survey
(o) Annual core variables
(p) Surveys to monitor impact of development projects

66. The sample design for Ethiopia is oriented towards estimating agricultural production and thus a member of Farmers' Association or a holding is used as a unit. Since there are households with more than one member, design becomes non-self-weighting for other socio-economic surveys.

67. The survey design for Kenya is formulated to make the probability of selection of a cluster of an EA within each district constant. Instead of selecting EAs in the first stage and clusters in the second stage, clusters might as well be selected directly with equal probability to obtain replicated samples or simple random samples. However though these designs can become more efficient by eliminating one stage of sampling, more than one cluster can be selected from the same EAs.

68. It appears that except Ethiopia these countries will use rotating samples either in each survey or after each survey to reduce a component of non-sampling errors. The designs of Cameroon and Kenya are formulated especially to make them self-weighting whereas those of the other two countries are not certain to be self-weighting. The advantage of having a self-weighting design is that data processing and integration become easier. The summary or advance tables can be prepared conveniently either in the field or at regional offices without using a computer. On the other hand the self-weighting design is not as flexible as a non-self-weighting design in allocating units in the later stages and in distribution of work-load to interviewers. Many self-weighting designs end up as non-self-weighting due to defective frame and undercoverage of units. If countries can solve problems arising out of those in getting summary or advance tables and preparation of computer programmes then they might as well opt for non-self-weighting designs. However many countries find data processing a big constraint and therefore Cameroon and Kenya try to develop self-weighting designs by making the probability of selection of penultimate units constant and selecting a constant fraction of ultimate units. And integration is an added attraction for self-weighting design.

69. Sampling units and sample sizes (number of households) are described in Table 1. Some countries use dwelling unit or holding (member of FA) as an ultimate sampling unit instead of households for various reasons. But, since information on households is of primary interest in many socio-economic surveys, it will be more appropriate to select households in the ultimate stage. Sample sizes appear to be large enough for general surveys, however, larger sizes would be necessary in some countries for demographic and labour force surveys.

70. Sample size differs in the surveys of those countries depending on the objectives of the survey programme, design and resources at their disposal. It appears that all these countries envisage to integrate the surveys. With the sample design and size and core variables planned to be used in those surveys, integration or linkage of surveys is possible. The question is the level of integration which will depend on:

(a) common core items, concepts and definitions in questionnaires in surveys,
(b) sample design and common sampling units in surveys,
(c) control of non-sampling errors to get consistent data in the different rounds,
(d) capability in on line data storage, retrieving, collecting, matching and linking, i.e., establishment of data base, and
(e) capability to analyse those data.

These countries will hopefully try to address these points and some results are expected to come out after a few years experience in this area.

71. The list of field staff indicates that those in Ethiopia and Kenya are strong, that in Cameroon is smaller but that in Botswana is the smallest. Field staff in many countries except those in rural areas of Ethiopia are mobile. Kenya has requested for an expert in Systems Analysis and some short term consultants. Botswana and Ethiopia have requested international experts in many areas. Cameroon is currently receiving technical advisory service from ECA and may probably request some short missions by the household survey specialists and data processing experts of UN and other agencies. But that country has no provision in the budget for consultants and experts.

72. These countries have drawn up programme to train the local staff at different levels. Some professionals and semi-professionals will be sent abroad or to the Statistical Training Centres in Africa and some junior and field staff will be trained locally. After completion of training and current survey programmes these countries are expected to be capable of continuing the survey programme with little or no external assistance. Ethiopia has requested external assistance for training centre.

73. Cost of the survey programme is the highest in Ethriopia and Kenya but the largest burden, i.e., more than 80 per cent of the cost will be borne by the Governments of those countries. On the other hand Cameroon and Botswana do not have such an expensive programme but the Governments would share the cost of the programme about equally with the external assistance.

74. Data processing is an area in which most of these countries are weak although some countries have their own staff. Ethiopia and Kenya would probably find data processing a bottleneck since many surveys are planned to be conducted every year. Special effort should be made to prepare editing and tabulation plans and programmes for each survey well in advance to ease the big burden in data processing.

75. Cartographic teams are established in those countries except Botswana to prepare and update census EA and segment maps. If a master sample is used their work will be confined to that sample only. Printing is not a big problem in these countries although some statistical departments do not have their own printer.

76. One of the biggest constrainst in these countries is transportation for the field staff to move from one place to another and for the supervisors to check the interviewers and to send the forms. Lack of transportation is responsible for non-sampling errors in many surveys particularly long-term repeated surveys. Therefore, if these countries cannot acquire the necessary vehicles it will be difficult to carry out the surveys successfully. Data processing is another constraint which has been addressed to in the above paragraph. small countries like Botswana face problem of recruiting staff.

VII. CONCLUSION

77. The common sample design used by these four countries is stratified multi-stage sampling with probability proportional to size. It is hoped that the estimates will be of the Horvitz-Thompson type to make it robust as explained by Godambe and Thompson.

78. At this initial stage of the survey programme, the assessment of the survey design in terms of sampling and non-sampling errors, cost, level of estimation and integration of surveys is too early to be realistic. This will be possible after a few years of survey experience in those countries.

79. This paper deals with the survey designs and their related aspects of four African countries which have joined AHSCP. Though there are other African countries participating in the AHSCP, the necessary information are not available to cover them in this paper.

REFERENCES

1. Republic of Botswana - Continuous Household Integrated Programme of Surveys (CHIPS), Project Proposal (January 1981).

2. Continuous Household Integrated Programme of Surveys (CHIPS) CHIPS SAMPLE DESIGN.

3. United Republic of Cameroon - Proposals for Household Survey Capability Programme.

4. Report of the Mission to the United Republic of Cameroon (28 February - 7 March 1981) by Myint Tin.

5. Proposal for setting up National Integrated Household Survey Programme (NIHSCP) in Ethiopia, prepared by the Central Statistical Office, Addis Ababa, Ethiopia, 1 January 1980.

6. National Household Survey Capability Programme Project Proposal for Ethiopia.

7. National Sample Survey and Evaluation Programme 1980/84, prepared by the Central Bureau of Statistics, Kenya, October 1980.

8. Republic of Kenya, Proposals for a Household Survey Capability Programme.

9. The National Household Survey Capability Programme, PROSPECTUS UNITED NATIONS, New York, 1980.

10. Godambe, V.P. and Thompson, M.E. (1977). "Robust Near Optimal Estimation in Survey Practice". Proceedings of the 41st Session of the International Statistical Institute, Volume XLVII - Book 3 - pp 129-146.