

BULGARIAN INTEGRATED HOUSEHOLD SURVEY - 1995
SURVEY ORGANIZATION AND IMPLEMENTATION

1) Description of the Project

1.1 Goals and Objectives

*The principal objective of the household data collection is to obtain a nationally representative household data set that contains detailed information on a variety of characteristics. This will allow a detailed analysis of the standard of living of the Bulgarian population in general and of the poor in particular, since it will allow to link information on income and expenditure to other household characteristics like health and education for example.

1.2 Organizational Structure

*Management, organization and field work is carried out by Gallup International in Sofia.

*The World Bank Office in Bulgaria will provide the necessary liaison, while the technical assistance will be furnished by World Bank staff and consultants provided from Headquarters.

1.3 Management

*The office of Gallup in Sofia under the direction of Dr. Zhivko Geogiev is in charge of the overall management of the survey.

*The co-ordination between the survey managers in Sofia and the field supervisors is assured by a layer of 9 main managers/supervisors in the 9 regional centers. They are directly responsible for the training of the other field supervisors and interviewers scattered in the 28 ex provincial districts centers. They are also responsible for the verification and implementation of the sampling frame, the data entry and the overall quality of the data. If the data is rejected they are ultimately responsible for having it corrected and sent to the field offices if necessary.

*The actual field operations are decentralized into the 28 field locations in the ex provincial centers. They contain at least one supervisor/interviewer. The actual number of interviewer in each of the centers will depend on the actual number of clusters that are selected in each of the centers.

*The list of survey management includes:

- Overall manager - Day to day manager and quality controller
- Data manager
- Regional Supervisors

2) Instrument

2.1 Coding and flow

*Except in a few cases the questionnaire contains all the codes necessary to be completed.

*All the codes are located under the questions they refer to, when possible, and in code boxes if the list is too long and they refer to more than one question. Code boxes are placed on the same or

adjacent page, only in the case of employment codes the interviewer is referred to a specific page at the end of the questionnaire.

* The flow of questions and skip patterns has been designed carefully to facilitate the interview process.

*The analysis of pilot test helped to verify the adequacy of coding and the flow of skip patterns - In general more coding is preferable to less, it is always possible to recompile the codes at the analytical stage, while re-coding afterwards can be a very tedious or impossible exercise.

Valid entries:

0 - used for all expenditures equal to zero

Valid blanks - for the all the non applicable values

Don't Know - Coded as NES

Refusals - Coded as OTC

2.2 Expenditure Section

*Particular care went into the design of the expenditure section given the current problem of approximately 100% inflation experienced in the past 12 months.

*The second peculiarity is the existence of lumpy purchases of food commodities that are believed to be in short supply, like vegetable oil and sugar; or canned foods that are received by relatives.

*To solve some of these problems, the expenditure section is based on a variable recall method designed to gather information on the monthly or weekly basis. The first set of questions deals with the actual consumption of food commodities, broken down by sources, regardless of the time of purchase. A second question deals with actual or estimated current prices. The third set of questions gathers the actual weekly or monthly expenditures made by the household during the past month.

* In each one of the subsections, the household is first asked if a specific item has been purchased or consumed in the past month. In case of affirmative answer then the other questions are asked.

*The units of measure of each item are specified next to the item. Units of measure have been standardized, if there are some commodities that might be specified in other units, conversion codes can be supplied in the field manual or by supervisors with flash cards.

2.3 Time of completion and number of visits

* The questionnaire has been designed to be administered in one session, which can be accomplished in one or more visits. Usually one visit may be sufficient, unless the interviewer has to gather specific information of a member of the household that was absent at the time of the first visit.

*It also possible that the interviewer has to go back to the same household to complete some of the unanswered questions or to correct inconsistencies.

*The expected time for the completion of the interview process will vary. During the pre-test the actual time taken to complete the interview varied from 1 and half to two hours in the 2 to 3 members households with some interviewers taking over 3 hours in some cases. But, as it is in all surveys of this type, the length of the interview in the pilot is longer than usual because of lack of experience. In fact the time necessary to learn and master an

integrated survey is quite long and usually takes several questionnaires before interviewers get accustomed to the flow of the questions and the location of the codes.

2.4 Community questionnaire

*The community questionnaire will not be administered in the first wave of the survey. Some information on the availability of services are available in the questionnaire. If needed, a community questionnaire can be administered in the following waves.

3. Sampling

3.1 Sample size

*The desired sample size is approximately 2,500 households over the whole national territory of Bulgaria, excluding replaced households.

General Census information - Last done in 1992

*In Bulgaria there are approximately 40,000 Statistical Sectors (SS). Each one contains approximately 80 households, giving a total of 2.9 million households. The average household size is 3 and the total population is approximately 9 million (8,459,763 at the last count in December 1993).

3.2 Raising factor

*Assuming the same household size of 3, a sample of 2,500 households will have approximately 7,500 individuals. This brings the raising factor (the ratio between total population and sampled individuals) to approximately 1,200.

3.3 Sample Selection

*A fixed number of households (5) is collected in each SS. The number of SS to be selected will be 500 ($5 \times 500 = 2,500$). Each cluster is selected with probability proportional to the number of households in each cluster from a subsample of 4,000 available representative clusters.

*The general sample methodology has been designed to insure that all the households of different size and in all regions and town have the same probability to be selected. The actual sampling of the households will be done in two stages: a) In the first stage the SS are Selected with Probability Proportional to the Size. The listing has been organized in the following order: (i) the list of the 28 regions, (ii) Cities, towns and villages; (iii) each city and town is listed by size; b) In the second stage the households to be interviewed are selected with Equal Probability, provided that they have been listed according to their size (all households of one person first, all the households of two people second and so on)

3.4 Replacement

*One of the major objective of the integrated household survey is

to reduce the number of refusal to a minimum level of under 5 percent.

* Each refusal has to be verified by the team supervisor and each substitution has to be authorized by the field supervisor as well.

*All refusals and their reasons will have to be recorded. A copy of the cover page of the questionnaire will contain all the information relative to the household location, time of the visit and reason for the replacement.

4) Data collection and organization

4.1 Role of central office

* The central office in Sofia will be responsible for the overall co-ordination and they will make sure that the work is carried out in a similar way in all areas.

4.2 Role of Regional offices

*Regional offices located in the 9 regional capitals, will provide the bases for the regional supervisors, who are responsible for the quality of the data in their region.

*The data entry will be actually done in 4 of these offices.

4.3 Location and responsibilities of 28 Field offices

* Each of the field office will receive a specific number of SS to complete. This number will depend on the proximity to the regional office and the number of interviewers assigned to that location.

* The total requirement for the number and type of SS in the 28 regional offices is contained in the table in the appendix.

* The Supervisors in these locations will be responsible for the field supervision of the data collection in the area and verification of their quality.

4.4 Structure of survey teams

* Usually this type of surveys are carried out by a small number of teams. Each team is composed of one supervisor and at most 5 interviewers and one data entry operator. This is to increase the proficiency of the interviewers with the questionnaire and to allow the most supervision for each team.

*The total number of interviewers usually depends on number of questionnaires in such a way that each interviewer should do between 50 to 100 questionnaires. This is necessary to assure proficiency that can be acquired only after some time.

*In this survey most of the field supervision responsibilities have been assigned to the supervisors in the 28 field offices. The number of interviewers assigned to each of the 28 offices will depend on the actual number of SS selected in each one (refer to the table of SS). It also possible that in small regions, the supervisors will fill in as interviewers.

*Since the SS contain only 5 households, each interviewers will be assigned a number of SS to complete and will work independently in each one of them.

*The overall quality of the field procedures and data quality is

overseen by the 9 regional supervisors.

4.5 Collection requirements

*Each interviewer can complete approximately 2 questionnaires a day.

*Each interviewer can be responsible for 10 Questionnaires a week or in other words 2 Statistical Sectors a week.

*The whole survey can be carried out in approximately 5 weeks.

4.6 Publicity and approach to the household

*Each interviewer will be carrying an identification card with photo, name and telephone number of the organization.

*The interviewer will be responsible to explain verbally the purpose of the survey and to convince the household to participate to the study.

4.7 Training

*The 9 regional supervisors were trained by Gallup managers and the WB consultants during the testing of the questionnaire. It was extremely important to involve them in the project and increase their contribution towards the preparation of a good questionnaire.

*The introduction of the questionnaire to the district supervisors and several interviewers took place on the Feb. 19, 20 and 21 in Varna, Plodviv and Sofia.

*In the next step, interviewers will fill 1 to 2 questionnaires and will be monitored by the 9 regional supervisors.

*The 9 regional supervisors will then come to Sofia to discuss and analyze the questionnaires completed with the help of the Survey managers, and will have the data entered in the computer. They will also receive training for the supervision of the quality of the listing and the selection of the households.

4.9 Survey Material

* The necessary survey material will be furnished to the field teams (Pencils, erasers, calculators, flash light, etc.)

5) Data Entry and Data Management

5.1 Data entry program

*The data entry program has been developed by a World Bank consultant (Beatriz Godoy)

*The design of the data entry screen follows the lay-out of the questionnaire.

*All the codes are included in the program. Valid entries are:

Values

Zeroes - Especially in case of expenditure

Blanks - From non applicable following skip patterns

Refusal -

Do not Know -

*The variables in the program have been specified with the question numbers. Labels and codes are specified on the files that control the data entry screens.

*Several checks are performed at the time of the data entry: (i) Ranges checks for quantitative variables; (ii) Screen checks (intra-record) to verify the consistency of variables included in the same screen; and (iii) Global (Inter-record checks) to verify the consistency across variables included in different screens.

5.2 Teams of operators

*The supervision of the data entry operators will be done by the regional supervisors.
 *The training of the data entry operators will take place in Sofia, with the use of the questionnaire collected during the pilot and the training exercise at the time when the 9 regional supervisors come to Sofia for their briefing.
 *The expected daily workload of the data entry is expected to be approximated 10 a day. This figures has to be verified at the time of the training of the data entry operators.
 *The data entry operators will need to process approximately 100 questionnaires a day. The total number of data entry operators will be equal to approximately 10 (100/10) [Figures to be verified at time of training]
 *The location and number of data entry operators will depend on the workload in each data entry center.
 *The data entry will be conducted at the same time of the data collection.

5.3 Data correction

*A printout, with detailed analysis of the data, is prepared for each questionnaire.
 *In case of a simple keypunching error, the operator is responsible for the correction.
 *In case of cross-check errors, the supervisor can attempt to correct the inconsistency. In case he/she is not able to do so, the questionnaire should go back to the supervisor of the location that collected the data.

5.4 Data quality

*Overall control is performed by the daily manager.
 *The detailed quality control is performed by the regional supervisors.

APPENDIX

Regions and Provincial distribution of Hhs

Region	Reg Name	Prov	Prov Name	Interviewed*			Population
				HHs	Res	All	
1	Sofia City	21	Sofia City	385	972	985	1,190,123
2	Bourgas	2	Bourgas	125	347	354	440,372
2	Bourgas	19	Sliven	62	190	199	234,785
2	Bourgas	28	Yambol	55	167	172	176,552
3	Varna	3	Varna	140	433	446	464,945
3	Varna	24	Dobrich	61	197	204	232,780
3	Varna	27	Shumen	60	179	181	220,320
4	Lovech	4	Veliko Tarnovo	98	285	288	318,252

4	Lovech	7	Gabrovo	48	125	130	161,987
4	Lovech	10	Lovech	55	127	140	190,262
4	Lovech	14	Pleven	95	271	285	346,614
5	Montana	5	Vidin	37	91	92	151,636
5	Montana	6	Vratsa	80	212	219	270,679
5	Montana	11	Montana	65	150	152	208,198
6	Plodviv	12	Pazardjik	85	239	255	326,123
6	Plodviv	15	Plodviv	205	618	648	734,495
6	Plodviv	20	Slolyan	50	134	143	159,752
7	Russe	16	Razgrad	45	120	125	167,410
7	Russe	17	Russe	80	202	208	288,702
7	Russe	18	Sillistra	45	131	139	161,063
7	Russe	25	Targovishte	39	105	109	151,339
8	Sofia Region	1	Blagoevgrad	95	311	315	351,637
8	Sofia Region	9	Kyustedil	59	167	168	181,347
8	Sofia Region	13	Pernik	55	132	132	163,307
8	Sofia Region	22	Sofia Region	90	293	295	289,962
9	Haskovo	8	Kardjali	55	192	197	213,806
9	Haskovo	23	Stata Zagora	114	343	363	397,339
9	Haskovo	26	Haskovo	85	249	255	295,503
			TOTAL	2468	6982	7199	8,489,290

*HH = Number of households interviewed

Res = Number of respondents interviewed

All = Number of individuals in the households