

PN-ACN-960

**BANGLADESH HOUSEHOLD
COPING STRATEGY SURVEY
(November 1998-December 1999)**

Data Documentation

FMRSP-IFPRI

October 2001

**International Food Policy Research Institute
2033 K Street, N.W.
Washington, D.C. 20006-1002**

www.ifpri.org

1. Introduction

The original objective for collecting this detailed household data set arises from the concern about food security of the rural households and non-availability of job opportunities during the flood and in the period following the flood, and to suggest policy measures to improve household food security in a sustainable way. The lessons from the response of the people and the government to the flood will not only be important in case of another disaster, but also to improve the food security of poor and landless households in time of stress. It may be noted in fact that the period following the flood is traditionally a period of food scarcity in most areas of Bangladesh. In Bangla it is the month of Katric, which loosely speaking means dreadful month.

There has been virtually no comprehensive quantitative information available for the period immediately after the flood apart from a few studies conducted by some organizations, who have either collected qualitative information or concentrated on their project sites where they are working or concentrated on a few key outcome variables. To that end, it was essential to document information on household composition, education, current and past employment, level of expenditure and other important characteristics of the households like assets, quality of housing, and health outcomes. Thus, the data collected contain all dimensions of household characteristics. This makes this a very good data set to analyze the impact of the flood, as well to perform many other welfare and poverty studies.

The data collected covered approximately 750 households in rural areas at three points in time over a period of a year between November 1998 and December 1999. The actual data collection took place in Nov-Dec 1998, Apr-May 1999 and Nov-Dec 1999.

This document is organized as follows. In section 2 there is a brief description of the survey instrument. In section 3, the sample design. The data entry and data management are discussed in section 5. In section in 6 and 7 the original and supplemental data files are outlined. Notes to potential users are presented in section 8.

2. Survey Instruments - Questionnaire

2.1 *Summary description of the household questionnaire*

The data collected includes variables necessary to describe and model several

dimensions of the household in sufficient detail. Some of the sections of the questionnaire were particularly designed to determine the particular exposure of the household to the flood and strategies used to mitigate its impact. Particularly important are the sections relative to the ownership and disposal of assets and the borrowing strategies. In addition, because one of the goals of the survey was to obtain data on the flood that devastated the country in the fall of 1998. Therefore in several section of the questionnaire the information is requested for the current and each of the past three years.

Box 1. Summary Description of the Content of the Questionnaire

1.	Household information	Sec A1 contains the usual information on the roster, like age, gender, civil status, time of absence from the household and so on. In addition it asks if the individual send or receive money for support.
2.	Education	Sec A2. concerns with the questions on education level for all individuals age 6 and older, dropout, and if any development programs running with the school.
3.	Status and history of employment, job search, training and public works	<p>The employment section is limited to all household members age 10 and over.</p> <p>In sec B1 there are questions relative to the labor participation, the main type of work and the reason for not participating.</p> <p>Sec B2 contains questions relative to the job search strategy and the attitude towards accepting a job (willingness to relocate and minimum wage). Also contains the history of employment held before the current employment.</p>

		Sec B3 accommodates information on Training and public works. Here the main questions relate to the number of weeks spent in public works and job training for each year since 1995.
4	Dependent job, Permanent and daily labor	Sec B 4 contains information on primary and secondary dependent job: Type of job, industry, time allocated, type of contract, salary and benefits on three different times frame.
5	Casual jobs, daily labor	Sec B 5 informs on time spent, tasks, wage rates etc. of causal jobs for three time periods.
6	Non-ag self employment, Business Activities	Cottage Activities, non-agri self-employment's information for three different time periods.
7	Agricultural activity, access to agricultural land, production and allocation of production	Sec C is dedicated to the agricultural production Availability of agricultural land, agricultural assets and livestock. In sec 6.1 the number of weeks worked during the past year and the hours worked last week are reported. Details on access (for each of the past four years) and type and acquisitions of agricultural land (orchard, pastures and cropland) are reported here.
8	Fishing activity, access to agricultural land, agriculture assets and livestock	Sec D, Sec E and Sec F.
9	Social assistance,	The sections on Social assistance and social benefits

	availability of benefits	are the central part of the questionnaire. In sec G, there is the level and the number of months several benefits received, currently and in the last three years.
10	Housing, ownership of real estate assets, Household furniture and durable	The section on housing contains the usual questions relative to the quality of the dwelling and the rent paid together with other monthly expenses. Section H relatives to the household durable contains the number of items, the current value and the year of acquisition. Sec I Sec J
11	Credit	
12	Housing and Sanitation	Non food expenditures include regular non-food spending for the past month in sec O and occasional non-food spending occurred in the past 12 months in
13	Regular and occasional non-food spending	sec 11.2 (12 items).
14	Food expenditure and consumption	The food section (sec K, L, N) contains food items and meals consumed at home and away from home. For all the items that have been consumed during the last month quantities consumed from purchases, own production and received from other sources are listed

		along with the purchase value, if quantities are not known, and current price.
15	Health status	Health status includes type of disability and treatment for chronic illness (sec P) and treatment, cost and type of consultation for acute illness occurred in the past 4 weeks.
16	Anthropometry	Height and weight have been collected for all children below 100 years of age and all females between the age of 13 and 45.

The questionnaire is organized in several parts that can be grouped in 16 main sections and several subsections. A complete copy of the questionnaire is available upon request. A brief description of the content of the questionnaire organized by the main sections is presented in box 1.

3. Sample Design

3.1 General Description and Sample size

The goal of the sampling design was to select those areas, which would give a fair representation of the parts of the country affected by flood. In particular, for the in depth household survey we interviewed 757 household in seven flood-affected thanas.

The seven flood affected thanas were selected using three main criteria. First, we used the severity of flood as determined by the Water Board. They classified thanas to be not affected, moderately affected and severely affected, depending on the level and depth of the floodwater. Second, we used the percentage of poor people in the district in which the thana is located. Thanas with more than 70 percent of the population below the poverty line were classified as poor. Third and final, among the thanas included in each of the categories, we selected those thanas that have been included in other studies and that would give us a good regional and geographical balance throughout the six

administrative divisions of the country.

Households were randomly selected using multiple stages probability sampling technique¹. In the first stage, three Unions in each *Thana* were randomly selected. In the second stage, six villages were randomly selected from each union with probability proportional to the population in each village. Then, in each village two clusters (*paras*) were randomly selected using pre-assigned random numbers. Finally, three households were randomly selected in each cluster from a complete list of all households in the cluster (*paras*). As a result, we selected approximately six households per village, 36 per Union, 108 per thana for a final sample size of 757 households in 126 villages.

Three different instruments were used. A community questionnaire was used to collect information at the union level during the flood. A village level survey was conducted during November and December 1998 in 64 villages to collect information on rural labor market. A detailed household questionnaire was used to collect information on the pattern of household expenditure, the pattern of land use at the level of plot, and the participation to the rural labor market, the ownership and loss of assets, borrowing strategy and anthropometry. Several sections in the questionnaire contained retrospective questions on the situation during and before the flood.

The detailed household survey was administered at three different periods in time to capture the difference in labor participation and food security in the period following the flood and to understand the capabilities of recovering from the shock of the flood. The first round of data collection took place between the 3rd week of November to 3rd week of December 1998. The second round for the data collection was carried out between April

¹ In Sauria thana this was not done because we were using the random sample used by another IFPRI study.

and May 1999. And finally the third round of data collection was taken place in November 1999, exactly a year after the first round.

Sample Selection:

ZILA	UPZA	UNION	Village	N_Dist.	N_Thana	N_Union	N_Village	HHOLD	HH CODING		
									H. Thana	H.UP	H. Village
13	95	15	1631	Chandpur	Shahrasti	Chitasi	CHANDAL	210	7	15	1
13	95	15	3801	Chandpur	Shahrasti	Chitasi	KADRA	219	7	15	2
13	95	15	5431	Chandpur	Shahrasti	Chitasi	LEPSA	23	7	15	3
13	95	15	6191	Chandpur	Shahrasti	Chitasi	NARAHO	162	7	15	4
13	95	15	9181	Chandpur	Shahrasti	Chitasi	DAKSHIN TULA (BARA TULA)	152	7	15	5
13	95	15	9741	Chandpur	Shahrasti	Chitasi	UTTAR TULA (CHHOTA TULA)	77	7	15	6
13	95	65	1411	Chandpur	Shahrasti	Dakshin Meher	BISHARA	82	7	65	1
13	95	65	2221	Chandpur	Shahrasti	Dakshin Meher	DEBKARA	583	7	65	2
13	95	65	5111	Chandpur	Shahrasti	Dakshin Meher	KRISHNAPUR	95	7	65	3
13	95	65	6471	Chandpur	Shahrasti	Dakshin Meher	NIJ MEHER	977	7	65	4
13	95	65	6472	Chandpur	Shahrasti	Dakshin Meher	CHEKUTIHA	68	7	65	5
13	95	65	6571	Chandpur	Shahrasti	Dakshin Meher	NOAGAON	361	7	65	6
13	95	95	541	Chandpur	Shahrasti	Tamta	UTTAR BALASID	392	7	95	1
13	95	95	2441	Chandpur	Shahrasti	Tamta	DHUSUA	177	7	95	2
13	95	95	3531	Chandpur	Shahrasti	Tamta	ICHHAPUR (RAJAPUR)	432	7	95	3
13	95	95	7171	Chandpur	Shahrasti	Tamta	PARANPUR	313	7	95	4
13	95	95	9021	Chandpur	Shahrasti	Tamta	SONACHO	239	7	95	5
13	95	95	9731	Chandpur	Shahrasti	Tamta	UARUK	383	7	95	6

Table 1. Sampling Frame by Location

	Rural areas
Thanas	7
Unions	21
Villages	890
Para	
Number of households	757

3.2 *Non response and replacement procedure*

The original sample was done without replacement and the interviewers were instructed not to replace households that were not found or that refused to participate in the study. In the first round 757 households participated in the study. In the second round, 7 households either refused to be interviewed or were absent at the time of interview. And in the third round of the survey, there were 23 households who refused to be interviewed or were absent at the time of the survey.

4. Organization and Procedure of Fieldwork

The field data collection was contracted to DATA, a local consulting company lead by Zahidul Hassan Zihad, Wahidur Rahman Quabili and Md. Zobair. Part of the contract was funded by the FMRSP-IFPRI and part from the WB office in Dhaka.

4.1 *Design, translation, testing*

The questionnaire was design by Carlo del Ninno in collaboration with Paul Dorosh and other local collaborators from the FMRSP project and BIDS

The questionnaire was translated into Bangla by DATA. The quality of the translation was verified at the time of the field test when the answers coming from the filed were discussed and verified.

4.2 *Structure of the interview*

Separate male and female questionnaires were administered in the households.

The male questionnaire dealing with labor and agriculture section took about 2 hours and 45 minutes (on average) to complete. The female questionnaire, centered mostly on the food purchase, allocation and intake of food in the past 24 hours, took about 2 hours 30 minutes (on average) to complete

4.2 *Structure of survey teams*

A total of four survey teams were used for the data collection. Each team included a field supervisor, three male interviewers, three female interviewers and a driver. In total four supervisors and twenty-eight interviewers worked on the survey. A team of two completed a para

- Each interviewer completed on average a total of 54 questionnaires, ranging from a minimum of 1 to a maximum of 27 households.

The quality of the interviewers was very high. Most of the interviewers had a University degree. Most of them had been experienced widely in data collection with different international organizations.

4.3 *Training*

The majority of the training took place in DATA office and in the field. The training includes detailed discussions of the questionnaire over a period of five days, some practice interviews under the supervision of the trainers and discussions of the completed practice interviews. Given the small number of teams (only three) and the large number of interviews completed by each interviewer which on other hand ensure the good quality of data.

4.4 *Publicity and Remuneration to the households*

The fieldwork was conducted using a low key approach. The team supervisors visited the community and discussed the survey and its purpose with the local Government officials, elites etc.

A small gift was also given to the households that participated in the survey. The exact gifts given varied from time to time. Sometimes it was dish bowls, cash, soaps etc.

The reception to the interviewers has been very good. The participation of the

household was excellent. This attitude was reflected in the low number of actual refusal and minimal item non-response.

5. Data Entry / management

5.1 Data Entry Program and Verification

The data entry program was designed by Syed Abu Bakar Siddiq using IMPS, a data entry package developed by the US Census Bureau. The program was designed in such a way to follow the same layout of the questionnaire and allowed three types of data checks:

- a) Range checks,
- b) intra-record checks to verify inconsistencies pertinent to a particular section of the questionnaire and
- c) inter-record checks to determine inconsistencies between the different sections of the questionnaire.

Some detailed checks were performed to verify the wage level, food prices and also if the per capita quantities of food were inside some acceptable ranges. The program for the checking the first two types of errors was performed while the data were entered in the computer. The inter-record checks were performed for one or more households at one time, after the data were entered in the computers. (The original IMPS dictionary of variables and the program for error checking is available upon request).

5.2 Data Entry Operations

The screen layouts of the data entry and the error messages were in English. The data were entered in FMRSP-IFPRI office in Dhaka under direct supervision of Syed Abu Bakar Siddiq, by seven data entry operators that worked on two to three shifts a day. The data entered using IMPS are be stored in files that contain the data relative to one or more household questionnaires. In this case it was decided to group the households by the Union, where the data were collected. This facilitated the management of the data and the checking of the errors for the inter-record checks. The error report relative to each Union was given to the supervisor that worked in that area who was responsible for the .

Revisits to the households were not done on a routine base.

5.3 *Organization of Data for Analysis*

Once the data had been collected and checked with the data entry program the data files for the analysis were prepared. A special custom program, designed in Power Basic, was prepared to read the original IMPS files, to rearrange them according to the sections of the questionnaire, to add additional labeling information and to create the programs used by the statistical packages to import the data. While it is convenient to store the data at household level when the data are entered (facilitate running inter-record checks), it is difficult to use it in that fashion to analyze it. The original data files from the household are grouped in to files organized according to the sections of the questionnaire.

5.4 *Preliminary data cleaning*

The data cleaning was done at three points in time. First the data was checked at the time of the data entry, then was checked for each round with error checking programs written in STATA, then the data from the three rounds was checked to make sure that the data was consistent.

6. Organization of electronic data files - Original data files

6.1 *Description*

The electronic data files derived from the survey and prepared for analysis comprise a group of 42 hierarchical data files. The pattern of data files and names of the original files follow the same pattern and structure of the questionnaire. Some of the files are organized at the household level, some of them at the individual level and some of them at the food commodity level and so on.

Each file contains identification variables that allow for the merging and matching of the information to create new files that contain the variables needed for the analysis. A detailed list of the 42 original data files is in box 2.

Box 2. Original data files

<i>Num.</i>	<i>File Name</i>	<i>Description</i>
1	SEC00	sec10: Cover page
2	SECA1	sec11: Roster
3	SECA2	sec12: Education
4	SECB1	sec13: Labor – Main
5	SECB3	sec14: Public-works & training
6	SECB4	sec16: Job-Main dependent
7	SECB5	sec17: Labor – Casual
8	SECB6	sec18: Activity- Business & cottage
9	SECC1	sec19: Land-Owner & operation
10	SECC2	sec20: Land-Agricultural activities
11	SECC3	sec21: Land- Kitchen Gardens
12	SECC4	sec22: Land-Summary of agricultural production
13	SECD1	sec23: Pond-Fish culture
14	SECD2	sec24: Pond-utilization
15	SECD3	sec25: Pond-Production & inputs
16	SECE2	sec26: Livestock- Possession,Income"
17	SECE3	sec27: Livestock- expenditure(J-N '98)
18	SECE4	sec28: Livestock- expenditure(D-J '98)
19	SECF	sec29: Family labor-agriculture, livestock, fishing"
20	SECG1	sec30: Revenue-Social assistance
21	SECG2	sec31: Revenue-allocation
22	SECG3	sec32: Revenue-Social assistance (For round3 only)
23	SECG4	sec33: Revenue-allocation (For round3 only)
24	SECH1	sec34: Assets-Ownership
25	SECH2	sec35: Assets-Sales
26	SECI1	sec36: Credit-received
27	SECI2	sec37: Credit- use, repayment"
28	SECI3	sec38: Credit- non member Household
29	SECO	sec39: Non food expenditure
30	SECJ12	sec40: Housing & sanitation
31	SECJ3	sec41: Housing & sanitation
32	SECJ45	sec42: Housing & sanitation
33	SECJ78	sec43: Housing & sanitation

34	SECJ91	sec44: Housing & sanitation
35	SECK	sec45: Food expenditure
36	SECL1	sec46: Individual-Food allocation
37	SECL2	sec47: Individual-Food allocation
38	SECN	sec48: Eating habits
39	SECP1	sec49: Morbidity
40	SECP2	sec50: Morbidity
41	SECP3	sec51: Morbidity
42	SECQ	sec52: Anthropometry

6.2 *Documentation, Codebook and summary statistics*

The questionnaire contains most information about the data. The questions have been laid out clearly and the interviewers were asked to follow the questions literally. Most instructions are printed in the respective sections. Most of the codes are included in the box relative to the question. In a few cases they are reported in box in the same page where the questions are asked.

There has been an intensive use of skip patterns to facilitate the data collection and minimize the time spent filling the questionnaire. Skip patterns are represented by an arrow followed by the number which refers the next question to be asked or the next section (→8). In all the cases a skip pattern applies the data in the skipped questions will appear as missing. Truly missing values refer only to the questions that were not supposed to be skipped and that received no answers.

Simple description of files, variables and code labels along with the simple summary statistics are provided in text and MS Word format. Summary statistics are very useful to get a feeling of the data before they are used for analysis. They are also very important to verify that the data have been received in its integrity and has not been modified. Sometimes a small translation mistake from one format to another might change the nature of the data.

6.3 *Missing data and other Special codes, Data sets*

Missing values have been left blank on the forms and they are treated as "." (dot)

in most statistical packages. Refer to the statistical package used to get more details about their treatment. For example, in STATA they are not used to calculate sample statistics and they are assumed to be the largest numbers in the data set. In a very few cases special codes "99" or "98" have been used to highlight special situations. These are clearly marked in the questionnaire.

6.4 *Merging data from different data sets*

The data set is organized in several files of different levels of aggregation. Each household can be uniquely identified using the household identification variable *hhnum* (HouseHold Number). This is a variable made of 6 digits. The first digit refers to the code for Thana (administrative unit) of the household. The following two digits refer to the code of the Union (administrative unit) of the household. The fourth digit refers to the code for the Village of the household and finally the last two digits refer to the sequential identification of the household.

The *hhnum* is the only identifier for household level files. Individual level files have an individual code in addition to the household code: *pcode* (Individual id code). Of course all the individuals in the same households share the same household code. This allows to clearly identify the information relative to each individual in the data set. Similarly, other files of a different level of aggregation have additional identifiers. For example the food expenditure files have a unique code for each commodity that has been consumed by the households *fooditem* (code of food).

This method assures that the data are stored in the most efficient way and that the necessary information can easily be combined in the analysis to compare the existing variables and to create additional ones. Household level files can be joined together to combine variables from different files using the variable *hhnum*. Similarly, individual level files can be merged together to perform some analysis in which the information on age, education level and wage rate is needed. In this case the key variables would be *hhnum* and *pcode*. Of course it is also possible to add household level information to an individual level file or a commodity level file. The only caveat is to be careful about the keys that are used to sort and merge the data to make sure that the resulting file contains the data for the same individual, household etc.

7. Organization of electronic data files - Supplemental and Constructed data files

7.1 Description

There are three additional files that are part of the core data set. These files include information that is needed in order to be able to process the data properly. It includes a file with the information about the sample and the weights, a cross cutting file with information on location and other often used variables and a file that contain the calculated measure of welfare. These files are reported in box 3 and the summary statistics are available in appendix D.

Box 4. Additional Data Files

General

<i>Num.</i>	<i>File Name</i>	<i>Description</i>
1	HHSIZE	Hh size variables
2	HHMEMBER	Ind. Definition of Hh members
3	STRATA	Location, Flood Exposure, Land ownership

Expenditure

<i>Num.</i>	<i>File Name</i>	<i>Description</i>
1	CALORIE	Caloric and Protein conversion table
2	EXPF4	Food expenditure, individual items, calories etc.
3	EXPG1	Food expenditure, groups, calories etc.
4	EXPNF1	Non Food expenditure, individual items
5	EXPNF2	Non Food expenditure, group items
6	TOTHEXP	Total HH food and non food expenditure, calories and food shares

Health and Anthropometry

<i>Num.</i>	<i>File Name</i>	<i>Description</i>
1	ANTRO1	Individual characteristics and anthropometrics

Credit

<i>Num.</i>	<i>File Name</i>	<i>Description</i>
1	Credit1	History of borrowing by loan amount taken (in cash and kind), date loan is contracted, date of repayment of loan, duration of loan, repayment history, reason and source of credit, annual interest rate, amount outstanding etc.

The file STRATA contains basic village level and household information that is common across all files.

The file TOTHEXP contains the measure of total welfare, calculated as total aggregated expenditure that is made available to interested users. A detailed explanation of the methodology used to calculate total expenditure is contained in appendix XX.

8. Notes to Potential User

8.1 *Obtaining the data*

The data set is property of the FMRSP-IPPRI and the World Bank. Those interested in using the data should contact Dr. Carlo del Ninno at c.delninno@cgiar.org.

8.2 *Disclaimer*

The data has been collected using a specific sampling design and users should take that into account. This data set cannot be representative of the rural areas in Bangladesh, but only 7 thanas affected by the flood in the fall of 1998. In addition, the data can be regarded as representative only up to a certain level of desegregation that contains a minimum number of cases.

Several cleaning procedures have been performed. The actual forms have been checked in the field by the supervisors. The data entry program controlled for typing errors and logical errors. Additional programs have been run in Dhaka. Still, there are probably several outliers and inconsistencies in the data that have not been discovered or that have not been modified. In fact the principle utilized has always been to modify the data if the replacing value can be trusted to be true. This is the case for typing errors and column shifts. Ultimately, it is up to the person analyzing the data to decide a methodological strategy to deal with missing values and outliers.

8.3 *Data formats*

The data are available in STATA format