

GUYANA LIVING STANDARDS MEASUREMENT SURVEY

DATA ENTRY MANUAL

Guyana Bureau of Statistics
The World Bank
January 1993

OVERVIEW

Objectives of the Guyana Living Standards Measurement Survey (LSMS)

The primary objective of the project is to collect individual and household data which can be used to analyze relevant policy questions, particularly those relating to poverty, income distribution and the living standards of Guyanese households. The information gathered is intended to provide knowledge about the extent of poverty in the country, the impact of existing government and non-government agency benefits, and the appropriate design of future programs. The Guyana LSMS will include specific modules on Health, Education, Fertility, Migration, and Anthropometrics. The majority of the questions are asked for each member of the household, with precoded answers provided just below the question. Code lists are kept short and simple to minimize errors.

Methodology of the Survey

The LSMS will be implemented in conjunction with UNDP's Household Income and Expenditure Survey (HIES) and will draw data from one round (the third round) of HIES's 7000 household sample. Because each round represents a national sample, it is possible to administer the LSMS in one round only, covering approximately 1800 households in 616 enumeration districts. All households included in the HIES third round will also be included in the LSMS sample.

Once the HIES has been administered to a household, the interviewer will return to the household for a second (or third) visit to carry out the LSMS questionnaire. Many of the initial questions on the LSMS questionnaire will be copied directly from the HIES, eliminating the need to re-interview households.

Data Entry Considerations

Data will be entered directly from the questionnaire. Everything written on the questionnaire will be entered in the computer. Enumerators have been asked to write legibly, in capital letters, and using arabic numerals. In writing amounts and figures, each group of three figures should be separated with a decimal point, starting from the right. For instance, "one hundred thousand" should be written as "100.000" not as "100000" or "100 000". For questions to which the reply is a quantity or amount, only the numeral should be written in the appropriate box, without the unit of measurement. For example, if the reply is: "fifty dollars", "50" should be written, not "\$50."

Structure of the questionnaire

GLSMS will collect information on several different aspects of living standards. The questionnaire is organized accordingly in the following sections:

- Cover (household identification and survey control data),
- Household roster (basic demographic information for each member of the household),
- Health (to be asked of each person regardless of age),
- Education (to be asked of each person 3 years or older),
- Migration (for all household members living away from Guyana for more than 6 months),
- Fertility (data from one selected woman in each household and all the children she has given birth to), and
- Anthropometrics (weight and height of all children 59 months or younger).

Organization of the survey

The GLSMS will visit 1,800 households over a three-month period. These households are clustered in 616 locations --called "Enumeration Districts", or "EDs"-- distributed throughout the country. An average of 12 households are visited in each ED.

The survey data collection and data management process for each ED is organized as follows:

- (1) First, a team of interviewers visits all households in the ED and completes all questionnaire sections (except anthropometrics).
- (2) Questionnaires are revised by a survey supervisor and brought to GBS headquarters in Georgetown.
- (3) Data from all questionnaire sections in each household (except anthropometrics) are recorded in a personal computer by a data entry operator.

The computer controls that all data are correct and the operator marks in the questionnaire any errors or inconsistencies that are detected.

- (4) Interviewers and anthropometrists visit the ED a second time. While anthropometrists weigh and measure all children under age 5 in the survey households, interviewers re-ask any questions that the computer considered incorrect.
- (5) Questionnaires are brought back to GBS.
- (6) The operator enters into the computer all data in the anthropometric section of the ED questionnaires. At the same time, the operator corrects any inconsistencies in the rest of the sections, entering the correct answers to the questions that were re-asked.

This procedure ensures that the GLSMS data will be available for analysis shortly after the survey is conducted. The data will also be of a very good quality, because inconsistencies will be corrected by actual re-visits to the households.

Data entry operators play a crucial role in the process, being entirely responsible for many of the activities described above. To perform these tasks efficiently, they should be familiar with their computers, as well as with the GLSMS data entry program and certain data management standards to be followed carefully throughout the survey.

The rest of this manual is addressed to the GLSMS data entry operators and data managers. It describes how to use the data entry program and gives basic standards for the data entry stage of the GLSMS data management.

The reader is supposed to be familiar with the GLSMS questionnaire and to have already some basic experience with personal computers running under DOS. Specifically, knowledge of the following DOS commands is assumed:

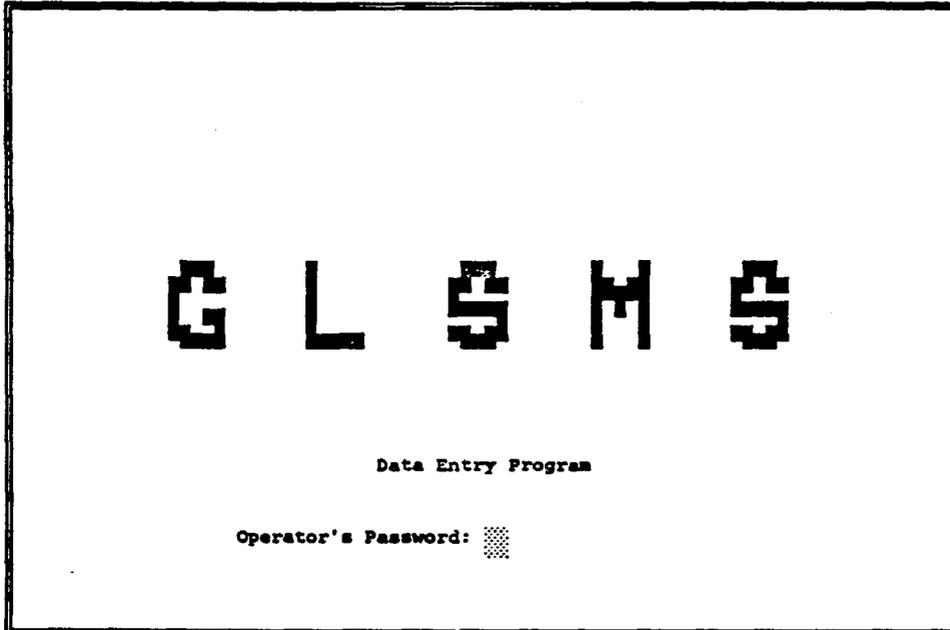
- FORMAT (to format 3½" and 5¼" diskettes),
- DISKCOPY (to copy diskettes),
- CHDIR (to navigate through disk directories), and
- XCOPY (to copy files and directories).

Using the data entry program

The data entry program is installed in directory C:\GLSMS of all the survey computers. From this directory the program can be started with the command:

```
C:\GLSMS>GLSMS [←]
```

The first screen will then immediately appear:



Each data entry operator is given a personal password. The system will wait for that password to be entered and check that it is one of the valid passwords for that computer. Another screen will then show up:

HOUSEHOLD IDENTIFICATION							
REGION (4)	SECTOR (6)	STRATUM NUMBER (7)	MAJOR DIVISION NUMBER (8)	ENUMERATION DISTRICT		SUB- SAMPLE NUMBER (11)	SAMPLE HOUSEHOLD NUMBER (12)
				NUMBER (9)	SERIAL No. (10)		
••••	••	••••	••••	••••	••••	••	••••

This screen is used to identify the household that will be entered. This is done through a series of numeric codes that appear in the cover page of each questionnaire (Region, Sector, Stratum Number, etc.). These codes are entered one after the other in data entry fields that show up on screen in the same order they appear in the cover.

The operator should take care to avoid making mistakes when identifying the household. If an incorrect code is entered, then information may accidentally be entered in the wrong file. If an incorrect code is entered, the operator should press the [Esc] key and start again.

Once all codes have been entered, the next screen will immediately appear. It shows the Main Menu of activities to be performed for a particular household:

HOUSEHOLD: [REDACTED]

----- MAIN MENU -----

- 1 Data entry
- 2 Display
- 3 Listing

- 4 Inter-record checks

- 5 Summary of all records entered
- 6 Complete listing of the household
- 7 Erase the household

Select an action: [REDACTED]

In order to enter data for a particular household, the operator should press 1 to select Data Entry. The next level is reached and the following menu is displayed:

HOUSEHOLD: [REDACTED]

----- QUESTIONNAIRES -----

- C COVER
- R ROSTER
- H HEALTH
- E EDUCATION
- M MIGRATION
- F FERTILITY
- A ANTHROPOMETRICS

Select the section: [REDACTED]

The operator selects a questionnaire section by typing in its initial letter. All selections, except "F" (for the Fertility section), will immediately conduct to a data entry screen. If the Fertility section is selected, the operator will be asked to select one of the two parts of this

The answers to most questions have been encoded with numeric values. Numeric fields are entered by either typing its digits until the end of the field is reached or by hitting [←] after the last digit if the number is shorter than the length of the field. In the later case, the system will automatically right-justify the number and pad it with zeroes on the left.

Alphabetic fields are entered in the same way, but they will not be justified by the system.

Some keys have special uses during data entry:

- The "left" and "right" arrow keys ([←] and [→]) can be used to move the cursor within a field, one position at a time. The [Home] and [End] keys will move to the beginning or to end of the field.
- The "up" and "down" arrow keys ([↑] and [↓]) and the "tab" and "backtab" keys ([→|] and [|←]) make it possible to move to other fields in the screens to make corrections.
- The [Ctrl] key, used in combination with the [End] key, fills with blanks all fields from the one where the cursor is located until the last field in the screen.
- The [F10] key can be used to indicate a "Don't know" answer to a question. These are marked by the interviewers as "DK" codes in the questionnaire
- The "Enter" key ([↵]) is used to record the data on the screen, once all fields have been entered.
- The [Esc] key will bring back the section menu, abandoning all the data entered so far in the screen. More generally, the [Esc] key will always bring the operator back to the prior screen, so that hitting this key repeatedly will eventually exit from the data entry program to DOS.

As data are entered in each screen field, the system checks if they are correct. If errors are detected, the system produces an audible warning ("beep"), takes the cursor to the beginning of the field and starts blinking and alternating the value entered with signs that indicate the

nature of the error:

- If an invalid code is entered (for instance, if a number "3" is entered for a question that can only be coded "1" or "2"), then the value will blink and alternate with question marks ("?").

the same will happen if letters are entered in a numeric field, or if an invalid date (like February 30th) is entered in a date field.

- If a figure below the lower limit of the numeric field is entered, then the low value will blink and alternate with arrows pointing down ("↓").
- If a figure above the upper limit of the numeric field is entered, then the high value will blink and alternate with arrows pointing up ("↑").

If there is an error in a field, the operator should look for the data for that field in the questionnaire and recheck the data entered. If the error is a result of incorrect data entry on the part of the operator, s/he should re-enter the data and continue.

If it is not the operator's mistake --i.e., if the data was entered exactly as shown on the questionnaire and the system indicates an error, the operator should press the "Enter" key ([↵]). On the last line of the screen the system will display the following blinking message:

Enter your password if sure: 

At that point, the operator should re-enter the password. The field thus corrected will start to blink and the cursor will move to the following field.

If there are other such errors, the operator should proceed in the same way.

Consistency checks

After all fields have been entered, the operator orders the recording of a screen by hitting the "Enter" ([↵]) key.

At that point, the system cross-checks between fields. If inconsistencies are detected between the fields, the system indicates them by sounding a "beep" and putting all inconsistent fields to blink.

The error message appears at the bottom of the screen and the cursor moves to the beginning of the first field.

The operator should examine the fields one by one, comparing them with the data on the questionnaire. If the error was due to incorrect data entry, s/he should correct the wrong fields and try to record the screen again. Otherwise, the operator should hit "Enter" ([↵]) to force the entry of the inconsistent screen. S/he will be asked to confirm with the password as explained before.

Documenting the inconsistencies

As explained above, when the operator detects inconsistencies in the questionnaire, s/he is supposed to force its entry as it is, rather than trying to "guess" what the correct value should have been.

The reason is that these inconsistencies will be corrected by the interviewers during the second visit to the household, as explained in the introduction of this manual.

Every time the operator detects such an error, s/he will document it on the questionnaire itself, encircling the wrong answers with a red pen. This will allow the supervisors and interviewers to locate them easily when in the field.

Posting of data

If the system does not detect any errors or inconsistencies in the screen, the system will play a lively tune and record the data immediately. If the operator had to force the entry of wrong data in the questionnaire, the system will play a slow tune instead.

In any case, the shape of the screen will change after the data are posted:

- For most data entry screens, all fields will become blank, to allow the operator to enter data for another line in the questionnaire. For instance, after the data in the roster for the head of the household (person number 01) are entered, the system will expect the operator to continue entering information for persons numbers 02, 03, etc.
- Other screens are supposed to be completed only once in each household. These are the questionnaire cover and the first part of the fertility section. After these sections are entered the system will automatically go back to the previous menu so that the operator can select the next section to be entered.

Deletion of erroneous records

To delete an erroneous record, the operator should select the "Display" activity (option 2) on the Main Menu and then bring up on the monitor the contents of section that includes this record.

There s/he should move the cursor to the record in question and press the [Del] key. The system will confirm the operation asking for the password, delete the record and show the newly amended section.

When finished, the operator should press the [Esc] key as usual to get back to the previous menu.

Printouts of the results of data entry

In addition to marking inconsistencies in the questionnaires, the operators will prepare a two-part printout for each questionnaire entered. This printout consists of: (a) a listing of all sections entered and (b) a listing of inter-record checks. The operator should produce this printout after the whole questionnaire has been entered.

To prepare the listing of all sections entered, the operator should select "Complete Listing of the Household" (option 6) from the Main Menu. To prepare the listing of inter-record checks the operator selects option 4 of the Main Menu.

After the system completes each listing (a process that may take a few seconds), it displays the following menu:

```

Your listing is in a disk file

Do you want to ...
1 See it in the screen
2 Print a hard copy

Select an action:

```

To produce the printouts, the operator should first prepare the printer (i.e., turn it on and make sure it has paper) and then select action "2" in the menu above.

Both printouts should be stored together with the questionnaire and submitted to the data entry manager.

The operator or the data entry manager may occasionally want to produce a partial listing of a particular section of the questionnaire. This can be done by selecting option 3 ("Listing") from the Main Menu and

then the desired section.

A typical printout of a listing for a section is as follows:

```

HOUSEHOLD 0330303003003103 -- [R] : HOUSEHOLD ROSTER

      I          R          R R R R R R R R R R R R
      D          0          0 0 0 0 0 0 0 1 1 1 1
      1          1          2 3 4 5 6 7 8 9 0 1 2 3
  
```

01	JOHN	1	1	40	1	1	2	2		08	2		01
02	MARY	2	2	35	1	1	2	2		08	2		08
03	SUE	5	2	18	1	1	2	1	02	05	1	02	05
04	PETTE	5	1	16	1	7	2	1	02	04	1	01	04
05	JANE	6	2	02	1	7	1	1	03	06	2		08

Some values in the printout (like for instance the answers to questions 9 and 12 for person number 03 in the example above) may appear encircled in black boxes. These correspond to those fields which were blinking on the screen at the time of data entry.

The listing of inter-record checks typically looks like this:

```

Household 0330303\003003\103: Consistency tests

---- Error number 1 :
Woman 02 did not answer Q29 in Health

---- Error number 2 :
Woman 03 did not answer Q29 in Health

---- Error number 3 :
Age of Person 05 in Q1 in Education
is larger than Age in roster

---- Error number 4 :
Age of Person 05 in Q5 in Education
is larger than Age in roster

4 errors detected in this household
  
```

The operator should revise this printout carefully, to make sure that the errors pointed out by the system are not a result of incomplete data entry. If this is the case, s/he should enter the missing parts and then carry out the inter-record checks again.

Data management

The basic unit for data management is the Enumeration Unit (ED). An ED is a cluster of households that are close to each other and are visited by a team of GLSMS interviewers at the same time. All questionnaires from an ED should be kept together at all times, both by the field supervisors and by the data manager and the data entry operators.

It is also important that data from all households in an ED are entered and stored always in the same data entry computer. Computers will be numbered for this purpose and the data manager should keep accurate records to ensure that this is done.

Enumeration Districts are identified through a series of six codes:

- Region (2 digits)
- Sector (1 digit)
- Stratum Number (2 digits)
- Major Division Number (2 digits)
- ED Number (3 digits)
- Serial Number (3 digits)

Within each ED, households are numbered from 101 to 112 and from 201 to 212. The first digit (1 or 2) indicates the Subsample code of the household.

Sixteen digits in total are thus needed to identify a GLSMS household (13 for the ED and 3 for the household within the ID). As the data entry program stores each household as a separate file in the disk, the use of such lengthy identifiers dictates that a subdirectory structure is needed to identify each file. All data subdirectories in each computer stem from directory C:\GLSMS\HH. Within this directory, several subdirectories contain the different "Major Divisions" (MDs) entered in this computer. The name of the subdirectory for an MD equals the first 7 digits of the ED number. Within each MD subdirectory, other subdirectories contain different EDs, each named with the remaining 6 digits. Finally, in each of these, different files, each with a three-digit name, contain the household data.

Inasmuch as the data entry programs manages this complex data structure automatically, the data entry operators don't need to be concerned with it. However, the data manager will need to perform frequent administrative operations, the most important being backing-up the survey

data regularly onto diskettes.

These operations can be performed with the aid of a special-purpose program called MANAGER, and are described in the rest of this manual:

Revising the contents of a data entry computer

To know the households that have been entered in a data entry computer, the manager should locate himself in directory GLSMS and type:

```
C:\GLSMS>MANAGER * ←
```

A screen will then appear with one line for each ED entered in the computer so far. If there are too many EDs in the computer, there may be no room for all of them in the screen, and the manager will have to scroll the list up or down using the [↑], [↓], [PgUp] and [PgDown] keys.

Each ED is identified by means of the six codes mentioned above. To the right of each ED, a series of "↑" symbols indicate which of the possible twelve households have been entered for each of the two subsamples in that ED. For example, the symbols

SUBSAMPLE 1	SUBSAMPLE 2
↑↑*↑↑↑*↑*↑*	↑*↑↑*.....↑

mean that households 101, 102, 104, 105, 106, 109, 201, 203, 204 and 212 have been entered.

The [Esc] key may be pressed to abandon the program at any time.

Backing-up data onto diskettes

To back-up data from an ED, the manager should mount a formatted, unprotected diskette with enough free space in the diskette drive and type:

C:\GLSMS>MANAGER A: ←

or

C:\GLSMS>MANAGER B: ←

depending on the name (A: or B:) of the drive.

A screen similar to a previous one will then appear, showing the contents of both the hard drive and the diskette. As before, a series of symbols indicate which households have been entered, but this time conventional symbols are used to indicate if a particular household is present in the hard drive (↑), in the diskette (↓) or in both (↑↓). For instance, the symbols

SUBSAMPLE 1	SUBSAMPLE 2
↑↑*↑↓↑↓*↑*↑*↑*↑	↑*↑↑*.....↑

mean that households 101, 102, 104, 203 and 204 are in the hard drive only, whereas households 105, 106, 109, 201 and 212 are both on the hard drive and on the diskette (which in this case probably contains an earlier back-up of the ED).

One ED is always outlined with a different color on the screen. The manager can decide which one by means of the [↑], [↓], [PgUp] and [PgDn] keys.

To back-up the outlined ED, the manager must instruct the system to copy all its households from the hard disk to the diskette. This is done pressing the [F1] key. After a few seconds, the system will complete the operation and re-draw the screen to present the new situation. In the example above, the new screen would show that after the back-up all households are now in the hard drive and in the diskette:

SUBSAMPLE 1	SUBSAMPLE 2
↑↓*↑↓↑↓*↑*↑*↑*↑	↑*↑↓*.....↑

The manager should indicate clearly on the label of each diskette the numbers of the ED and the date of the back-up.

Each ED should be backed-up immediately after data from all households are entered, both for the first and second visits to the households. It is also extremely important that at least two back-up diskettes are conserved for each ED at all times.

The MANAGER program imposes no restrictions to the quantity of EDs that can be backed-up on one diskette. For an efficient management, however, it is important that an orderly policy is followed in this regard. The same back-up diskette should never be used in two computers. Also, Major Divisions should be kept together as far as possible.

Restoring data from diskettes

The MANAGER program can also be used to restore ED data from the back-up diskettes to the computer. This may be necessary if data are accidentally lost from the disk.

To restore data from a diskette to the hard disk, the manager must outline the ED on the screen and then hit [F2]. The system will copy all household files from the diskette to the hard drive in a few seconds.