

1999/2000 HOUSEHOLD EXPENDITURE SURVEY PROCESSING MANUAL

TABLE OF CONTENTS

- 1. Purpose of processing**
- 2. General comments**
 - 2.1 Household and other identifiers
 - 2.2 Use of red pens, neatness
 - 2.3 Completeness of workload documents
- 3. Processing of HES-1**
 - 3.1 Sequence and completeness checks
 - 3.2 Coding of occupation and industry
 - 3.3 Recording of "period covered"
 - 3.4 Treatment of non-standard "period covered"
 - 3.5 Treatment of multiple answers to one question
 - 3.6 Checking duplication within HES-1
 - 3.7 Treatment of non-housing loans
 - 3.8 Commodity coding of selected items
 - 3.9 Checking consistency with HES-2 and HES-3
- 4. Processing of HES-2**
- 5. Processing of HES-3**
 - 5.1 Deletion of HES-1 commodity items from diary
 - 5.2 Distinguishing cash and non-cash items
 - 5.3 Commodity coding
 - 5.4 Checking values (validity, decimal place)
 - 5.5 Treatment of missing days in HES-3
- 6. Processing of HES-4**
 - 6.1 Checking completeness of questionnaire set for each household
 - 6.2 Treatment of missing days in HES-3

1. PURPOSE OF PROCESSING

The main purpose of processing in this survey is to convert the complex set of data as recorded by the interviewer into a standard and easily tabulated set of data on the computer. Rupee values will be converted into monthly equivalents, and all values will be checked for appropriateness and completeness.

Processing comprises both office processing and computer processing.

On the one hand, office processing will mainly deal with the following aspects:

- (a) Checking that each workload is complete;
- (b) Clerically checking each form for completeness, legibility and correct values;
- (c) Coding of commodities, industry and occupation;
- (d) Converting the values and non-standard periods in “last payment”/“period covered” questions into standard periods;
- (e) Checking consistency between forms for each household;
- (f) Capturing the data onto computer file for subsequent computer processing.
- (g) Resolving any edit failures detected during computer processing.

On the other hand, computer processing will mainly deal with the following aspects:

- (a) Checking values and codes for validity, sequence checks and consistency checks;
- (b) Converting the values associated with standard periods in “last payment”/ period covered” questions into standard monthly values;
- (c) Calculating selected derived data items for subsequent tabulation;
- (d) “Expanding” diary values to standard monthly values (to be done if two-week diaries are used or if incomplete monthly diaries are received).

This current manual deals only with the office processing aspects, and may be revised as the processing system is further developed. Office processors should be familiar with the “Interviewer’s Manual” and in particular with the concepts and definitions used in the survey.

2. GENERAL COMMENTS

2.1. HOUSEHOLD AND OTHER IDENTIFIERS

Each document in the survey must be uniquely identified. The household number should appear on each form. In addition, the HES-2 form should have the person number and the HES-3 should have the diary number (as well as the total number of diaries).

If these identifiers do not appear on the forms when they are received in the office, try to determine the correct number by looking at other forms in the workload set or by looking at householder name, etc. Inform the supervisor of the missing identifiers so that he can ensure that the interviewer is corrected on this point.

2.2. USE OF RED PENS, NEATNESS

You should only use red pen when writing on the forms. This is so that it will be possible to distinguish interviewer notes from office processing notes at some later time.

Always write words and numbers neatly and clearly. This will speed and simplify subsequent data entry and edit resolution.

2.3. COMPLETENESS OF WORKLOAD DOCUMENTS

Before you start looking at a particular household, you should first check that all the documents for the workload, which includes the household, have been given to you. There should be one HES-1 for each selected household, one HES-2 for each person aged 15 years or more and at least one diary for each household.

3. PROCESSING OF HES-1

3.1. SEQUENCE AND COMPLETENESS CHECKS

Check that all questions have been appropriately answered. Check that the SUMMARY at the foot of page 1 has been completed and that the numbers agree with the total household size and with the number of HES-2 forms completed.

Use “date of birth” to calculate age and see that this agrees with the count of “under 15” and “15 or more” as shown in the SUMMARY.

Check that “Education” and “Economic” status are only recorded for those aged 15 or more.

Check that “Occupation” and “Industry” are only recorded for those aged 15 or more and “Economic status” codes 0 to 6 inclusive.

Answers MUST appear for the following questions: 2, 3, 4, 5a-o, 6, 11a-b, 12a-c.

If YES, in 5c, 5f-j, 5l-m or 5o, then payments will normally be recorded. If there is no payment, then the interviewer should have explained the discrepancy in a note. For example, the business pays for the car, electricity, etc.

Question 6a is a major sequencing question. If code 1 (owner occupied) then 6b should be answered but not 6c or 6d. If codes 2, 3 or 4 (rented), then 6c should be answered but not 6b or 6d. Finally if code 5 (rent free) is answered then 6d should be answered but not 6b or 6c.

If regular payments, especially wage deductions and hire purchase payments, appear in the HES-1 then you should check the diary to make sure that these are not also recorded there. (As explained below, if they are recorded, then you should delete the entries in the diary with a single red line. This check is important because “deletable” diary items are easily overlooked).

Check that the interviewer has properly followed the instructions in sections 11 and 12 of the HES-1.

3.2 CODING OF OCCUPATION AND INDUSTRY

You should code all occupations and industries shown on page 1 of HES-1 for employed persons aged 15 years or more. The occupation and industry code lists are provided separately.

Write the occupation and industry codes in the space provided in the form.

If the person is shown as being employed (codes 0 through 6 in the “Economic status” column), but no description of occupation or industry has been recorded, then check if the “Economic status” code is correct. If the person is really employed, then the “not stated” code for these classifications will be “99”.

No codes are required if the person is aged under 15 years or if the person is not economically active (codes 8 and 9 in “Economic status” column).

3.3 RECORDING OF “PERIOD COVERED”

“Period covered” appears in several questions and will be used in computer processing to convert the “last payment” into a monthly equivalent.

The code for “period covered” should be recorded in two parts – the second part showing whether the period is in weeks (W), months (M) or years (Y), and the first part showing the number of such units of time.

Some examples are:

1Y = one year

1M = one month

2M = two months

3M = three months

1W = one week

2W = two weeks

4W = four weeks

Note that 4W (four weeks) is not the same as 1M (one month). There are 12 months to a year and 13 four weekly periods in a year, so the difference is about 8 percent. Please make sure that you record the correct code.

3.4. TREATMENT OF NON-STANDARD “PERIOD COVERED”

The computer system can only be expected to handle the normal cases. Sometimes, the respondent might report an unusual “period covered” such as 64 days. In this case, check to make sure that the answer appears correct and, if so, convert the period and associated value to a monthly basis prior to data entry.

Write the new value in the “last payment” space and “1M” in the “period covered” space, then delete the original entry with a single red line.

3.5. TREATMENT OF MULTIPLE ANSWERS TO ONE QUESTION

For most questions, there should be only one answer. Some questions (e.g. life insurance, wage deductions, and so on) may have more than one answer recorded but the computer data entry expects only one answer. More than one answer can even occur for questions such as electricity payments if the person pays for his own dwelling as well as for some other dwelling (such as for a parent).

If more than one answer is recorded for a particular question, carry out the following steps.

- (a) First check that the answers relate to the same commodity type. If not, then split the payments into groups of payments with each group relating to the same commodity. The following steps should be taken for each separate commodity group of payments.
- (b) If the question is on a “last payment/period covered” basis, then convert all the multiple answers for that commodity into a common period (such as one month). If the question is a “12 month” or a “3 month” basis then no conversion is required.
- (c) Add the answers to arrive at a total expenditure on that item and then record that total in the data entry space.
- (d) Put a single line in red through the components that you have converted and added up so that these components are not inadvertently entered into the computer file.

3.6. CHECKING DUPLICATION WITHIN HES-1

The HES-1 has been carefully designed to avoid duplication as much as possible. However, it is possible that regular expenditure paid by wage deductions (in HES-2), for example, may be duplicated with an item already recorded in the HES-1. For example, rent for housing will have been reported in 6c but may also be reported in HES-2 as a wage deduction by the employer.

If this occurs, neatly delete the wage deduction entry. Do not enter this information into the computer because this would result in duplicate expenditure.

Note that the regular payment of hire purchase should not be deleted in this way. A deposit may have been made during the reference period and should be retained in the appropriate section (last 12 months or last 3 months questions). There is no duplication between the deposit and the HP payment, and in fact the two items together will (on average over all households) reflect the total expenditure on the commodity.

3.7. TREATMENT OF NON-HOUSING LOANS

No data entry is required for expenditures recorded in 7j ("Non-housing loans"). As explained in the Interviewer's Manual, the loan from the bank, etc will have been spent in an outright purchase, which should have been recorded if it occurred in the reference period. On average over all households, this balances out and there would be duplication if the expenditure on repaying the loan was also to be included.

The data is collected so that the householder feels that all current expenditure is being covered. The data could also be used to calculate expenditure on loan interest, by assuming that a certain percentage of the repayment was on interest. However, this has not been recommended because the interest component of a loan repayment generally decreases as the loan is paid off.

3.8. COMMODITY CODING OF SELECTED ITEMS

Generally, the HES-1 items have been pre-coded with their commodity codes. However, some items in the HES-1 will need to be commodity coded, namely 8i ("hire purchase") and 8j ("non-housing loans").

As explained in the HES-3 section below, the prefix of "1" shown in the HES-1 can be ignored and all you need to do is to code the commodity with a 3-digit code using the commodity code list provided.

You may find it useful to also use the alphabetic-to-numeric index to facilitate this coding.

3.9. CHECKING CONSISTENCY WITH HES-2 AND HES-3

You should check the consistency between forms. For example, if a person is recorded in the HES-1 as being a wage earner, then you should expect wage/salary income in HES-2. Similar comments apply to employers and self-employed persons who should have business income.

You should note that it is possible for a wage earner to also have business income and for a businessman to earn wages in a second job.

If “own produce” has been reported in Section 12 of HES-1, then you should check for consumption of own produce in the diaries.

During this consistency check, you should account for diary “deletables” as described in section 5.1 below.

4. PROCESSING OF HES-2

The HES-2 form is very simple. There are basically only six new data items (relating to wage income, business income and other regular incomes).

Data on “sex” and “date of birth” will be added in data entry as a check to make sure that the person number shown on the HES-2 agrees with the person number from HES-1. It is recommended that you also check this consistency data entry in order to avoid computer edit failures.

In sections 2 and 3 of HES-2, the interviewer may have been required to clerically add and subtract various values in order to arrive at the income figures. Please check the arithmetic to make sure that no error has been made by the interviewer. It is easy to make a mistake while under pressure during the actual interview. Any errors should be corrected by you and the new values recorded.

5. PROCESSING OF HES-3

The most time consuming task in office processing for this survey will be the coding of commodities in the HES-3 diaries. This will also be the most important because estimates of all food costs and many other expenses will be derived from these diaries.

5.1. DELETION OF HES-1 COMMODITY ITEMS FROM DIARY

As explained in the Interviewer's Manual, survey estimates of expenditure will be derived from both the HES-1 and HES-3 forms. If we used all the data in HES-3 then we would be duplicating expenditure on the items which have already been covered in the HES-1. You must therefore delete with a single red line every transaction in the diary that relates to a commodity that has already been covered in the HES-1.

You must be very careful in carrying out this deletion because if you delete a transaction which should not be deleted, then the expenditure involved will be lost. If you do not delete an item which should be deleted, then a computer error may be indicated during computer processing.

In general, the commodity codes have been chosen so that codes for HES-1 items are separate from codes for other items. For example, "public telephone calls" are code 721 but "telephone bills" (which are covered in 7c-d of HES-1) are code 720.

An exception to this relates to "wage deductions" and "hire purchase repayments". Any expenditure in the diary relating to these two types should be deleted even if the expenditure is on a commodity that is not explicitly covered in the HES-1. This is because the "wage deductions" and "hire purchase payments" are explicitly covered in the HES-1 and HES-2.

For example, normally an item such as an "electric shaver" (which is not directly covered in the HES-1) should not be deleted from the diary. However, if it is paid for by hire purchase and the HP payment appears in the diary then the regular payment on this item is given the code 905 in HES-1 (question 8i) and the regular payment for HP becomes "deletable" in the diary.

5.2. DISTINGUISHING CASH AND NON-CASH ITEMS

When all the "deletable" items have been neatly crossed out in the diary, you can then code the remaining items. The form has been pre-printed so that:

- purchased items will have a prefix of "1" to their commodity code and
- items which have been consumed from own produce or been received free by the household have been given a prefix of "2" to their commodity code.

These two prefixes are so that we can later distinguish cash expenditure from non-cash consumption.

5.3. COMMODITY CODING

Except for the prefix (see 5.2), the codes for a commodity which has been purchased will be the same as the code for a commodity that has been consumed from own produce, etc. For example, bananas which have been purchased will have code “077” with a prefix of “1” (1077) and bananas which were home grown and consumed by the household will have a code of “077” with a prefix of “2” (2077).

You will quickly get to learn the codes for the most common expenditure items, but you will find the alphabetic-to-numeric code index to be most useful in determining codes for other items.

Commodity coding is a skilled function. You should use all the available information including the quantity, description and brand names, as well as the actual commodity description, when deciding what commodity code to allocate to the entry.

5.4. CHECKING VALUES (VALIDITY, DECIMAL PLACE)

You should make sure that the values recorded for each commodity seem reasonable. For example, an expenditure of about 5 rupees per kilo for rice is reasonable. If an expenditure of 500 rupees is shown for “rice” in the diary, you should check if a 100 kilo bag has been purchased (if so, this value should be accepted) or if the decimal place has been missed (in which case, you should amend the entry).

5.5. TREATMENT OF MISSING DAYS IN HES-3

If the diary is incomplete, the interviewer should have recorded the number of missing days in the HES-4.

This data will be used in computer processing to arrive at a “diary expansion factor” which will be equal to:

- 31 divided by (31 minus the number of missing days) if monthly diaries are used; or
- 14 divided by (14 minus the number of missing days) if two weekly diaries are used.

This process is quite straight-forward if only one diary (or one set of diaries) is used for the whole household or if all diary keepers in the household cease diary keeping on the same day.

However, it becomes complicated if some diary keepers stop maintaining their diaries before other diary keepers in the same household. In this case, the easiest solution is for you to go through the diaries of those who continued diary keeping after the others had stopped, and to delete all entries after the date on which the other diaries ceased. In this way all diaries in the same household will be reduced to a common period, and the “number of missing days”, (and hence the “diary expansion factor”) will be the same for all diaries in the same household. You may need to amend the “number of days missing from HES-3” in the HES-4 if you have reduced all diaries to a common period in this way.

6. PROCESSING OF HES-4

6.1. CHECKING COMPLETENESS OF QUESTIONNAIRE SET FOR EACH HOUSEHOLD

You should check that the information in the HES-4 (if any) on which forms have been completed agrees with the forms for that household. In other words, that the status of the HES-1, HES-2 and HES-3 forms have been correctly recorded. If not, make sure that you have the full set of forms for the household or amend the HES-4 entries as appropriate.

The HES-4 data should be captured onto the computer in all cases. However, there is little point in capturing the data for other forms (HES-1, HES-2 or HES-3) for the household if the data are unusable or if a successful substitute household was found.

You may need to discuss each HES-4 household with your supervisor in order to decide what to do with the other household forms. The following possibilities should be considered:

- (a) If the HES-1 and HES-2 forms are complete but the diaries are incomplete (but still usable after “diary expansion”), then the full set of household forms (including the HES-4) should be captured onto computer file;
- (b) If the HES-1 is complete and the HES-3’s are complete (or incomplete but still usable), but HES-2’s are incomplete, then the expenditure information is still useful and the set of household forms (including HES-4) should be captured to computer file. In this case, the income data will have to be recorded as “not stated”.
- (c) If the HES-1 is incomplete or the HES-3 forms are not usable, then the household data cannot be used and only HES-4 data should be entered.

In all cases, if a successful substitute household was fully enumerated, then the forms for non-respondent household should not be data entered (except for HES-4). If we entered the forms for the original selection as well as the substitute, then we have changed the sample and have interfered with the weighting scheme.

6.2. TREATMENT OF MISSING DAYS IN HES-3

As explained in 5.5 above, you may need to amend the “number of days missing from HES-3”.