

Data edits

National Baseline Household Survey Questionnaire, 2009

1. Pre- scanning and scanning edits

Pre scanning manual data checking:

- Supervisor checking and signing off in the field
- See fieldwork manual for priority checking

Data receiving/scanner feeding responsible at data processing centre:

- Check 1: Number of forms total per EA counted and protocolled
- Check 2: Staples removed before scanning
- Check 3: Scan 1 EA per "batch"
- Check 4: Re-staple, mark as scanned and store

Scanning verification on screen:

- Check 1: (must-be-filled-in-check)
If no codes for a1_state to a1_house, check TIFF file for text or writings outside box and put code based on text if possible – if not type 9, 99 or 999 (MISSING) to get past the check
- Check 2: (only-one-mark-allowed-check for all single response questions)
If more than one mark, check TIFF file and correct if possible - if not possible to decide on correction, type 9 or 99 (to signal to SPSS professional editor)
- Check 3: (valid-range-checks)
If outside range, verify TIFF on screen and be sure that what is written on the form is correctly interpreted (special focus on decimal errors and possible extra zeros given when writing SDGs). If errors identified then correct on screen, if not force the initial written value through without any changes. This will be dealt with in SPSS edits.

General rule for screen verifiers is "key what you see" i.e. no changes of data written on the form from the field during scanning.

2. Record Type transferred from scanning sequence

Only one record type will be defined per household and exported from the scanning. This is a semicolon separated record comprising all variables in ASCII format. During export, the file is split on 3-4 lines with CTRLF variables introduced on the scanning side because the un-split file (3659 variables total) exceeds the maximum record length that the current version of the software can export/import directly.

When importing the raw data ASCII file to SPSS, all variables for each household is merged back into one line i.e. one record per household, and the variables get names and formats according to the variable names agreed (see separate document).

1. Post scanning data processing basics

Data should be further processed when all households in a State is completed scanned and transferred to a state specific raw data file. Basically the programs (syntax files) should be executed sequentially on increasing syntax number where next syntax file builds on the outputs from the previous files. We recommend to run these files step by step and go through the log files after each job run. This gives better control than if we build a general macro that runs everything in one batch. On the other hand it requires know-how on SPSS and the functionality of the syntax files.

A float diagram should be annexed to the final version of this document for better understanding of the sequencing, in/out put of the programs.

The following overall rules/principles for editing are suggested:

1. Missing due to correct skipping, blank etc is set to "." and defined as SYSMIS (SPSS default)
2. True missing is defined as user MISSING and coded 9, 99, 999 depending of the initial variable length (SPSS can deal with 3 different codes for user-missing).
3. Consistency check on skip-patterns; if details are given, the skip question is adjusted to "yes" and vice versa. Changes of skip variable are flagged.
4. Suggested general rule for detecting variable outliers are if value $\geq 3 \times \text{stdv}$ then impute with median value. Changes of outliers are flagged **<Note: Outlier check may be improved when SPSS version 17 alternatives for data cleaning is further explored>**
5. With exception for the id-code for the enumerator (A1_7), MISSING is not accepted for section A1-A2, A4_1, A4_2 (number of forms used for the household) and B2-B4 (relationship, sex and age). A system for imputation & special checks for section A (look-up table from the sampling list) and section B (imputation) is required. Changes are flagged.
6. Flags are added as 1 digit variables at record level with the actual variable name + a leading "f_"
 - 0=no imputation
 - 1=logical from SYSMIS
 - 2=logical from invalid value
 - 4=imputation from SYSMIS
 - 5=imputation from outlier
 - 6=imputation from other invalid value
 - 7=imputation from manual check on TIFF files

SPSS editing of the raw data up to final production files per state is done in 2 main steps with possible additional third and fourth steps for more comprehensive imputation (3) and for prepared tabulation (4) to be discusses:

Step1: Restructuring and cleaning of raw data resulting in 8 temporary un-weighted work files

- a. Global file (one record per household) with corrected variables for section A and no duplicates (the structural edits)
- b. Person characteristics file (one record per person) for section B-F
- c. Housing and household characteristics file (one record per household) for section H-L + part of agriculture N1-3 and N6
- d. Food and beverages consumption file (one record per commodity) for section M1-3
- e. Non food purchases (one record per commodity) for section M4-5
- f. Crop plot module (one record per plot) for section N4
- g. Crop harvest module (one record per crop type) for section N5
- h. Household income module (one record per household) for section O (CBS only)

Step 2: Final production files (a-h) with weights, derived/grouped variables and variable labels

a-h. Final weighted and labelled production files

Possible additional products/steps:

Step 3: Alternative additional step with more complete imputation of variables from the step 2 files

Step 4: Prepared tabulations of variables from the step 2 file

Two additional data files (look-up tables) are required for the post scanning data edits:

1. A look-up table (nhbs_lookup1) with information from the sampling frame (state – boma names&codes, ea code, urban rural code and range of questionnaire serial numbers used in the ea.
2. A conversion table with the link between units of measure (heap, baskets, cups etc), and food-commodity in kg (nhbs_lookup2).

4.1 WORKFILE (a) (section A) file name: TMPWORK[state#]_ID

4.1.1 Questionnaire module: A0

Syntax file name: nbhs_01_import

Input:

Output: nbhs09_raw_state#.dat and nbhs09_raw_state#.dat

PURPOSE:

Raw data import and storage

1. Store initial ASCII file as 'nbhs09_raw_state#.dat
2. Import ASCII file raw data
3. Restructure raw data to one line per record
4. Allocate variable names and formats and delete the CTRLF variables
5. Export and store SPSS raw data file as 'nbhs09_raw_state#.sav'
6. Back up the State folder with raw data files on external media

4.1.2 Questionnaire module: A1-2

Syntax file name: nbhs_02_check A1-A2

Input: nbhs09_raw_state#.dat, nbhs09_lookup1.dat 1)

Output: tmp_1A_state#"

1) uhbs09_lookup1:

PURPOSE:

Create data file with new id variables "geo_code" and "household" cleaned data for variables in questionnaire module A1-A2.

Variable name: num_n

Question: ..

Universe: All forms scanned

Valid codes:

Edits & imputations

If num_1 <> num_(2-n) LIST and do manual corrections looking for possible mixing of pages from different forms in the same set. Manual check of TIFF files.

Variable name: a1_state, a1_county, a1_payam, a1_boma, a1_ea

Question: ..

Universe: All forms scanned

Valid codes: from "nbhs_lookup1" (table from sampling frame)

Edits & imputations:

IF a1_state = SYSMIS THEN set a1_state=99

IF a1_county=SYSMIS THEN set a1_county=99

IF a1_payam=SYSMIS THEN set a1_payam=99

IF a1_boma=SYSMIS THEN set a1_boma=999

IF a1_ea=SYSMIS THEN set a1_ea=999

IF a1_house=SYSMIS THEN set a1_house=999

COMPUTE variable "geo_code" as 12 digit code (state – ea)

IF geo_code <> "nbhs_lookup1". Flag and LIST, check TIFF and correct geo_code manually

LOOP geo_code check until all valid

Check and correct "a1_state" by matching with final "geo_code" and/or "nbhs_lookup1". Flag changes

Variable name: a1_house

Question: ..

Universe: All forms scanned

Valid codes:

Edits & imputations:

Sort by "geo_code" and "a1_house"

If a1_house is SYSMIS, find manually and fill in correct code (there may be duplicates see A4,1-2)

If no valid a1_house can be found, allocate artificial household number starting per ea with 800, 801 etc. COMPUTE variable "household" as an 15 digit unique identifier based on "geo-code" and corrected "a1_house".

Variable name: a1_worker

Question: ..

Universe: All forms scanned

Valid codes:

Edits & imputations:

If a1_worker is SYSMIS, and there is a duplicate "household", copy a1_worker from duplicates else set a1_worker = 999

Variable name: a1_urbrur

Question: ..

Universe: All forms scanned

Valid codes: from "nbhs_lookup1"

Edits & imputations:

Check match for a1_urbrur with LOOKUP table "nbhs_lookup1". If a1_urbrur is wrong or SYSMIS, then replace with LOOKUP table value. Flag.

Variable name: a1_north, a1_east

Question: ..

Universe: All forms scanned (for SSCSE only)

Valid codes: from ea maps

Edits & imputations:

If coordinate(s) are SYSMIS and there is a duplicate on "household", replace SYSMIS with coordinates from duplicate. Flag. Else if SYSMIS, add centre coordinates for the ea (from ea maps). If coordinates $\geq 3 \times \text{stdv of ea MEAN}$, the replace coordinate(s) with centre coordinates from ea-map. Flag.

4.1.3 Questionnaire module: A3-A4

Syntax file name: nbhs_03_check A3-A4

Input: tmp_1A_state#

4.2 Output: tmp_2A_state# (all variables) + TMPWORK[state#]_ID (only variables from A)

PURPOSE:

Clean the person file module A3-A4.

Completeness and consistency checks. Imputation. Flagging of changes.

Variable name: a3_yes1

Question: Did the household live in this county 12 months ago?

Universe: All forms scanned

Valid codes: 1, 2

Edits & imputations:

If a3_yes1=1 or SYSMIS and (a3_state >0 and a3_county >0) and (a3_county ne a1_county) or (a3_county >1) THEN set a3_yes1=2

If a3_yes1=2 or a3_yes1=SYSMIS and a3_state and a3_county and a3_country = SYSMIS THEN set a3_yes1=1

Variable name: a3_state

Question: If no, where did the household live 12 months ago?

Universe: All forms scanned

Valid codes: SYSMIS, nhbs_lookup1

Edits & imputations:

If a3_state = SYSMIS or a3_state not valid and a3_yes1=2 and a3_country<2 THEN check TIFF for manual correction of a3_state, if no text set a3_state=99

If a3_state >0 and a3_yes1=1 THEN set a3_state=SYSMIS

Variable name: a3_county

Question: If no, where did the household live 12 months ago?

Universe: All forms scanned

Valid codes: SYSMIS, nhbs_lookup1

Edits & imputations:

If a3_county=SYSMIS or a3_county not valid and a3_yes1=2 THEN check TIFF for manual correction of a3_county, if no text set a3_county=99

If a3_county>0 and a3_yes1=1 THEN set a3_county=SYSMIS

Variable name: a3_name, a3_country

Question: If no, where did the household live 12 months ago?

Universe: All forms scanned

Valid codes:

Edits & imputations:

If a3_name > 0 THEN check a3_country manually and compare with the written name:

1= Sudan

2= Other African country

3= All other countries

After manual check on text box and corresponding correction/coding:

If a3_country=2 or 3 THEN set a3_yes1=2 and a3_state = SYSMIS and a3_county=SYSMIS

If a3_country =1 THEN set a3_name="Sudan" and possible SYSMIS on a3_state and a3_county =999

Variable name: a3_yes2

Question: Does this household usually migrate during the year to look for work, raise livestock or for other reasons?

Universe: All forms scanned

Valid codes: 1,2

Edits & imputations:

If a3_yes2 ne 1 or 2 THEN set a3_yes2=9

Variable name: a4_totform, a4_thisform

Question: Total number of forms used for this household / Of which this is form number

Universe: All forms scanned

Valid codes:

Edits & imputations:

Must be filled in

Run duplicate check on "household" and sort duplicates on top of file

Identify and delete "true duplicates" (i.e. if a1-a4 and b1-8 and h1-10 and m2 and n1 is the same)

Loop duplicate test and be sure that all "true duplicates" are deleted

If ">12-member-household-duplicate" check a4_totform and a4_thisform and number them correctly. Important that head of household is on the first numbered form in households with more than one form used (more than 12 persons) and that ALL housing, health and economic information is on the last numbered form in the household (or at least not partially distributed over several forms). The "tot form and "this form" must be correctly numbered according to this order.

If no duplicate and SYSMIS a4_totform THEN set a4_totform=1.
If no duplicate and SYSMIS a4_thisform THEN set a4_thisform=1.

If no duplicate and a4_totform >1 THEN LIST and check manually
If no duplicate and a4_thisform >1 THEN LIST and check manually

Flag changes of a4_totform and a4_thisform

Variable name: a4 day to minute

Question: interview time

Universe: All forms scanned

Valid codes:

Edits & imputations:

If a4_month1(2) ne 4 or 5, THEN set a4_month1 and a4_month2 = 05 (CBS) and 04 (SSCCSE).

Set a4_year and a4_year2=09

If a4_day/ a4_hour/ a4_minute = SYSMIS, THEN set 99

Variable name: a4_check

Question:

Universe: All forms scanned

Valid codes: 1

Edits & imputations:

If a4_check = SYSMIS then set a4_ckeck=9

Run SPSS MVA or similar test on all variables for corrected module A printout report and store temporary work file

4.3 WORKFILE (b) Person file (section B-F) File name: TMPWORK[state#]_B

PURPOSE:

Transpose and clean a person file (one record per person) with corrected variables from module B-F. Completeness and consistency checks. Imputation. Flagging of changes.

EDITS/RESTRUCTURE:

- 1 Transpose all variables B1-F3
- 2 Keep selected key variables from section A1-4 on each person record: #date, #time, num_1, tiff_(1 – 8), a1_state, geo_code, household, a3_urbrur, a1_worker, all A2, all A3, a4_totform, a4_thisform, a4_month and a4_year
- 3 If no information filled in for module B to F, delete the whole record
- 4 If no information in module B, but some information I C-F (and vice versa), flag record and look/correct for possible “erroneous switching of records/columns”.
- 5 Redo 3-4 until only valid records are left
- 6 Sort by geo-code (from a1) and a4_thisform and renumber household members into computed new variable “member” with head of household as member number 1 in each household.
7. Compute new variable “tot_members” and allocate total number of household members to this variable

4.3.1 Questionnaire module: B

Syntax file name:	nbhs_04_check B
Input:	tmp_2A_state#
Output:	tmp_B_state#

PURPOSE:

Transpose person file

Clean the person file module B1-B8.

Completeness and consistency checks. Imputation. Flagging of changes.

Variable name: b2_(1-12)

Question: What is [Name's] relationship to head of household?

Universe: All persons belonging to the household

Valid codes: 1, 2...12

Edits & imputations:

If no head of household marked or head < 12 years and a person older than 12 years exists, set oldest person in the household as head. Renumber "member". Flag

Else If 2 or more head of household, set oldest person to head and others to code 11 (other relative). Renumber "member". Flag.

Else If b2_(1-12) = SYSMIS and B3 or B4 >0, THEN set b2_(1-12)=11. Renumber "member". Flag

Variable name: b3_(1-12)

Question: Is [Name's] male or female?

Universe: All persons belonging to the household

Valid codes: 1,2

Edits & imputations:

If head = male then spouse=female (and vice versa). Flag.

If only one daughter/son and this is = male, and male or SYSMIS for spouse of son/daughter, THEN set spouse of daughter/son=female (and vice versa). Same logic for Parent (if 2 persons) (b2_8). Flag

Else If SYSMIS, impute sex based on probability per State per subgroup of relation to head b2. Flag.

Variable name: b41_(1-12) and b42_(1-12)

Question: What is [Name's] age in completed years?/ If less than 5 years; What is [Name's] age in completed months?

Universe: All persons belonging to the household (b42 for persons less than 5 years)

Valid codes: 0-95

Edits & imputations

If age > 95 write 95. Flag.

0-5 years check:

If months = 0-59, set age = according to months. Flag

Else If months > 59, set months to SYSMIS. Flag

Else If months SYMIS and age <5 set months according to age (0,12, 24, 36, 48). Flag

Check consistency with relationship (b2), marital status (b5), highest level of education (c7) and work (D).

Head of household age checks:

if SYSMIS head age or head age < age of spouse and age of oldest spouse >12, set age=age of oldest spouse + MEDIAN difference of age between head and spouse in the State. Flag

else if SYSMIS head age and age of child ne SYSMIS set head age= age of oldest child+15

else if SYSMIS head age and age of female parent ne SYSMIS, set age = mothers age-15

else if head age = SYSMIS set age = MEDIAN age of head within State.

Check consistency with marital status (b5), highest level of education (c7) and work (D).

Other members age checks:

If age = SYSMIS set age = MEDIAN within State of subgroups of relation to head (b2). Flag
Check consistency with relationship (b2), marital status (b5), highest level of education (c7) and work (D).

Variable name: b5_(1-12)

Question: What is [Name's] marital status?

Universe: All persons belonging to the household age 12 years or older

Valid codes: 1-4 and 9 (no response)

Edits & imputations

For all persons if age < 12, set b5 = SYSMIS

Check head marital status towards b2 (spouse) if spouse, then head marital status and spouse marital status=2 and vice versa.

If only 1 daughter/son (b2_3) then set possible spouse of son/daughter b2_4= 2 married (and vice versa)

If other persons with age ≥ 12, and marital status = SYSMIS, THEN set status=9

Variable name: b6_(1-12)

Question: During the past 12 months, how many months did [Name] live in this household?

Universe: All persons belonging to the household

Valid codes: 0-12 and 99 (no response)

Edits & imputations

If b6 > 12 and b6 < 99, set b6=12. Flag

If b6 = SYSMIS then b6= 99

Variable name: b7_(1-12) and b8_(1-12)

Question: Is biological father (mother) of [Name] living in this household?

Universe: All persons belonging to the household

Valid codes: 1-4 and 9 (no response)

Edits & imputations

If b5 age ≥ 70 and SYSMIS b7(b8), set b7(b8)=3 (dead). Flag

Else, if SYSMIS then set b7(b8) = 9

Run SPSS MVA or similar test on all variables for corrected module B printout report and store temporary work file

4.3.2 Questionnaire module: C

Syntax file name: nbhs_05_check C

Input: tmp_B_state#

Output: tmp_C_state#

PURPOSE:

Clean the person file module C1-C11.

Completeness and consistency checks. Imputation. Flagging of changes.

For persons ≥ 6 years old

If b4_1 age < 6 set c1 – c11 = SYSMIS

Variable name: c1_(1-12)

Question: Can [Name] read and write with understanding a simple sentence in any language?

Universe: All persons 6 years and above

Valid codes: 1, 2 and 9 (no response)

Edits & imputations

if c1=SYSMIS and c5>1 or c6>2 or c7>2 set c1=1. Flag

if c1=9 (double entry from scanning) check TIFF and decide 1 or 2 manually

LOOP until correct

if c1=SYSMIS and c5, c6 and c7= SYSMIS set c1=9. Flag

Check that only valid codes are in

Variable name: c2_(1-12)

Question: Has [Name] ever attended school?

Universe: All persons 6 years and above

Valid codes: 1, 2 and 9 (no response)

Edits & imputations

If c2=1 and c5-c7=SYSMIS set c2=2. Flag

If c2=2 or c2=SYSMIS and c5 or c6 or c7 ne SYSMIS, set c2=1. Flag

If c2=9 (double entry from scanning) check TIFF manually and decide on 1 or 2

LOOP until correct

If c2=SYSMIS and c5-c7=SYSMIS set c2=9

Check that only valid codes are in

Variable name: c3_(1-12)

Question: Is [Name] currently attending school?

Universe: All persons 6 years and above

Valid codes: 1, 2 and 9 (no response)

Edits & imputations

if c3=1 and c5=SYSMIS set c3=2. Flag

if c3=2 or c3=SYSMIS and c5>0 set c3=1. Flag

If c3=9 (double entry from scanning) check TIFF manually and decide on 1 or 2

LOOP until correct

if c3=SYSMIS and c5=SYSMIS set c3=9

check that only valid codes are in

Variable name: c4_(1-12/1-7 i.e. multiple response allowed)

Question: If [Name] is less than 16 years old: Why is [Name] currently not attending school?

Universe: All persons 6-15 years

Valid codes: 1 and 9 (no reponse) multiple entry

Edits & imputations

if c4_(1-7)=SYSMIS and c3=2 then set c4_(1-7)=9

check that only valid codes are in

Variable name: c5_(1-12)

Question: For those currently attending school: What is the grade and level that [Name] is attending?

Universe: All persons 6 years and above

Valid codes: 1-17 and 99 no response

Edits & imputations

if c5= (1-17) and c3=2, set c3=1

If c5=99 (double entry from scanning) check TIFF manually and decide on 1 to 17

LOOP until correct

If c5=SYSMIS and c3=1 then set c5=99

If c3=2 set c5=SYSMIS

Check that only valid codes are in

Variable name: c6_(1-12)

Question: For those currently attending school: What is the grade and level that [Name] was attending previous school year?

Universe: All persons 6 years and above

Valid codes: 1-17 and 99 (no response)

Edits & imputations

If c6=99 (double entry from scanning) check TIFF manually and decide on 1 to 17

LOOP until correct

If c6=SYSMIS and c3=1 and c5=99 then set c6=99

If c3=2 set c6=SYSMIS

Check that only valid codes are in

Variable name: c7_(1-12)

Question: For those currently attending and/or previously attending school: What is the highest level that [Name] has completed?

Universe: All persons 6 years and above

Valid codes: 1-16 and 99 (no response)

Edits & imputations

If c7=99 (double entry from scanning) check TIFF manually and decide on 1 to 16

LOOP until correct

If c7= SYSMIS and c2=1 set c7=99

if c2=2 then set c7 = SYSMIS

Check that only valid codes are in

Variable name: c8_(1-12)

Question: Has [Name] ever attended any sort of vocational training?

Universe: All persons 6 years and above

Valid codes: 1, 2 and 9 (no response)

Edits & imputations

if c8=1 or c8=SYSMIS and c9=SYSMIS, set c8=2. Flag

if c8=2 or c8=SYSMIS and c9(1-9) >0, set c8=1. Flag.

If c8=9 (double entry from scanning) check TIFF manually and decide on 1 to 2

LOOP until correct

Check that only valid codes are in

Variable name: c9_(1-12/1-9 i.e. multiple response allowed)

Question: If yes, what type of skills did [Name] acquire?

Universe: All persons 6 years and above

Valid codes: 1 and 9 (no response) multiple entry

Edits & imputations

If c9(1-9) = SYSMIS and c8=1 then set c9(1-9)=9

Check that only valid codes are in

Variable name: c10_(1-12)

Question: How many months did [Name's] vocational training last?

Universe: All persons 6 years and above

Valid codes:

Edits & imputations

if c10=SYSMIS and c8=1, set c10=99. Flag

if c10>0 and c10 ne 99 and c8=2 set c10=SYSMIS

Variable name: c11_(1-12)

Question: How many months did [Name's] vocational training last?

Universe: All persons 6 years and above

Valid codes:

Edits & imputations

if c11=SYSMIS and c8=1 set c11=999. Flag

if c11>0 and c8=2 set c11=SYSMIS. Flag

COMPUTE new variable "voc_cost" with 1 decimal = c11/c10 (if c11 and c10 > 0 and ne 99 or 999

Range check if voc_cost >= 3 stdv of MEAN cost in State, replace MEDIAN voc_cost. LIST. Flag.

Adjust c11 and c10 accordingly. LIST. Flag

Run SPSS MVA or similar test on all variables for corrected module C printout report and store temporary work file

4.3.3 Questionnaire module: D

Syntax file name: nbhs_06_check D

Input: tmp_C_state#

Output: tmp_D_state#

PURPOSE:

Clean the person file module D1_D14.

Completeness and consistency checks. Imputation. Flagging of changes.

For persons >=10 years old

If b4_1 age < 10 set d1 – d11 = SYSMIS

Variable name: d1_(1-12)

Question: During last 7 days, did [Name] work at least one hour for pay (or without pay), profit in kind or for family business?

Universe: All persons 10 years and above

Valid codes: 1,2 and 9 (no response)

Edits & imputations

If d1=9 (double entry from scanning) check TIFF manually and decide on 1 to 2

LOOP until correct

if d1=1 and (d6-d10 and d12=SYSMIS) set d1=2. Flag

if d1=2 or d1=SYSMIS and (d6 or d7 or d8 or d9 or d10 or d12>0) set d1=1. Flag

if d1=SYSMIS and d6-d10 and d12=SYSMIS set d1=9

if d1=1 set d2-d5=SYSMIS

Variable name: d2_(1-12)

Question: [Name] did not work during last 7 days, but have job to go back to?

Universe: All persons 10 years and above

Valid codes: 1,2 and 9 (no response)

Edits & imputations

If d2=9 (double entry from scanning) check TIFF manually and decide on 1 to 2

LOOP until correct

if d2=1 and (d6-d10 and d12=SYSMIS) set d2=2. Flag

if d2=2 or d2=SYSMIS and (d6 or d7 or d8 or d9 or d10 or d12>0) set d2=1. Flag

if d2=SYSMIS and (d6-d10 and d12=SYSMIS) set d2=9

if d2=1 set d3-d5=SYSMIS

Variable name: d3_(1-12)

Question: [Name] did not work during last 7 days, but have worked before and is available for work

Universe: All persons 10 years and above

Valid codes: 1,2 and 9 (no response)

Edits & imputations

If d3=9 (double entry from scanning) check TIFF manually and decide on 1 to 2

LOOP until correct

if d3=1 and (d6=d10 and d12=SYSMIS) set d3=2. Flag

if d3=2 or d3=SYSMIS and (d6 or d7 or d8 or d9 or d10 or d12>0) set d3=1. Flag

if d3=SYSMIS and (d6=d10 and d12=SYSMIS) set d3=9

if d3=1 set d4-d5=SYSMIS

Variable name: d4_(1-12)

Question: [Name] did not work before, and is not seeking work

Universe: All persons 10 years and above

Valid codes: 1,2 and 9 (no response)

Edits & imputations

If d4=9 (double entry from scanning) check TIFF manually and decide on 1 to 2

LOOP until correct

if d4=1 and (d6=d10 and d12=SYSMIS) set d4=2. Flag

if d4=2 or d4=SYSMIS and (d6 or d7 or d8 or d9 or d10 or d12>0) set d4=1. Flag

if d4=SYSMIS and (d6=d10 and d12=SYSMIS) set d4=9

if d4=1 set d5=SYSMIS

Variable name: d5_(1-12)

Question: [Name] have never worked before, but is seeking work?

Universe: All persons 10 years and above

Valid codes: 1,2 and 9 (no response)

Edits & imputations

If d4=2 and d5 <> 1, set d5=1. Flag

Variable name: d6_(1-12)

Question: For those who worked or have worked before(yes in d1-d3): What was [Name's] main activity of work last 7 days or las days work?

Universe: All persons 10 years and above

Valid codes: 1-24 and 99 (no response)

Edits & imputations

If d1=1 or d2=1 or d3=1

If d6=99 (double entry from scanning) check TIFF manually and decide on 1 to 24

LOOP until correct

if d6=SYSMIS and d12>0 and d12 ne 99 set d6=d12. Flag

if d6=SYSMIS and d12=SYSMIS, set d6=99

if d1 =2 and d2=2 and d3=2 then set d6=SYSMIS

Variable name: d7_(1-12)

Question: Did [Name] receive wages, salaries or other values in cash or kind or in other values for the work done during the last 7 days?

Universe: All persons 10 years and above

Valid codes: 1,2 and 9 (no response)

Edits & imputations

If d1=1 or d2=1 or d3=1

If d7=9 (double entry from scanning) check TIFF manually and decide on 1 to 2. flag

LOOP until correct

if d7= SYSMIS and d8=SYSMIS and d9=SYSMIS set d7=9

if d7=1 and d8=0 and d9=0 set d7=2. Flag

if d7=2 and (d8>0 or d9>0) set d7=1. Flag

if d1 <> 1 and d2 <> 1 and d3 <> 1 THEN set d7=SYSMIS

Variable name: d8_(1-12)

Question: What was the value of [Name's] last payment (cash or kind)? (or if not yet received payment) What is the value of the payment that [Name] expect to receive (cash or kind)?

Universe: All persons 10 years and above

Valid codes:

Edits & imputations

If d1=1 or d2=1 or d3=1

if d8=SYSMIS and d7=1 set d8=999

if d1 <> 1 and d2 <> and d3 <> 1 THEN set d8=SYSMIS

Variable name: d9_(1-12)

Question: How many days did [Name] work for the payment just reported(d8)? (or if not yet received payment) How many days does [Name] expect to work for the payment just reported?

Universe: All persons 10 years and above

Valid codes:

Edits & imputations

If d1=1 or d2=1 or d3=1

If d9= SYSMIS and d7=1 set d9= 999

if d1 <> 1 and d2 <> 1 and d3 <> 1 set d9= SYSMIS

COMPUTE new variable "payday" with 1 decimal = d8/d9 (if d8 and d9 > 0 and ne 999

Range check if payday >= 3 x stdv of MEAN in State, replace with MEDIAN payday. LIST. Flag.

Adjust d8 and d9 accordingly. LIST. Flag

Variable name: d10_(1-12)

Question: For those who worked last 7 days or have worked before (from d1,d2,d3): What was [Name's] main employment status?

Universe: All persons 10 years and above

Valid codes: 1-5 and 9 (no response)

Edits & imputations

If d1=1 or d2=1 or d3=1

If d10 =9 (double entry from scanning) check TIFF manually and decide on 1 to 5. flag

LOOP until correct

if d10= SYSMIS, set d10=9

if d1 <> 1 or d2 <> 1 or d3 <> 1 set d10= SYSMIS

Check that only valid codes are in

Variable name: d11_(1-12)

Question: For those who did not work before and were not seeking work (from d4): Why did [Name] not seek work (mark only one)

Universe: All persons 10 years and above

Valid codes: 1-7, 9 (no response)

Edits & imputations

If d4=1

If d11 =9 (double entry from scanning) check TIFF manually and decide on 1 to 5. flag

LOOP until correct

if d11= SYSMIS and d4=1 set d11=9

if d4 <> 1 set d11= SYSMIS.

Check that only valid codes are in

Variable name: d12_(1-12)

Question: For those who worked last 7 days or have worked before (from d1, d2, d3): What was [Name's] main activity of work during the last 12 months (ISIC-4)

Universe: All persons 10 years and above

Valid codes: 1-24

Edits & imputations

If d1=1 or d2=1 or d3=1

If d12 =99 (double entry from scanning) check TIFF manually and decide on 1 to 24. flag

LOOP until correct

if d12= SYSMIS set d12=99

if d1 <> 1 or d2 <> 1 or d3 <> 1 set d12= SYSMIS

Variable name: d13_(1-12)

Question: For the last month (30 days) - What was the value of [Name's] last payment (cash or kind)? (or if not yet received payment) What is the value of the payment that [Name] expect to receive (cash or kind)?

Universe: All persons 10 years and above

Valid codes:

Edits & imputations

If d1=1 or d2=1 or d3=1

if d13= SYSMIS set d13=999

if d1 <> 1 or d2 <> 1 or d3 <> 1 set d13= SYSMIS

Variable name: d14_(1-12)

Question: How many weeks did [Name] work for the payment just reported(d14)?

Universe: All persons 10 years and above

Valid codes:

Edits & imputations

If d1=1 or d2=1 or d3=1

if d14= SYSMIS set d14=999

if d1 <> 1 or d2 <> 1 or d3 <> 1 set d14= SYSMIS

COMPUTE new variable "payweek" with 1 decimal = d13/d14 (if d13 and d14 > 0 and ne 999
Range check if payweek >= 3 x stdv of MEAN in State, replace with MEDIAN payweek. LIST. Flag.
Adjust d13 and d14 accordingly. LIST. Flag
if d1 <> 1 or d2 <> 1 or d3 <> 1 set payweek = SYSMIS

Run SPSS MVA or similar test on all variables for corrected module D printout report and store temporary work file
--

4.3.4 Questionnaire module: E (ONLY FOR SSCCSE)

Syntax file name:	nbhs_07_check E
Input:	tmp_D_state#
Output:	tmp_E_state#

PURPOSE:

Clean the person file module E1-E3.

Completeness and consistency checks. Imputation. Flagging of changes.

For persons < 5 years old

If b4_1 age >= 5 set e1 – e3 = SYSMIS

Variable name: e3_(1-12)

Question: Result of measurement

Universe: All persons less than 5 years

Valid codes: 1-4 and 9 (no response)

Edits & imputations

If e3=9 (double entry from scanning) check TIFF manually and decide on 1 to 4. flag
LOOP until correct

if e3= SYSMIS or e3 >1 and (e1>0 or e2 >0) set e3=1

if e3=1 and e1= SYSMIS and e2 SYSMIS, set e3=9

if e3=2-4 and e1 >0 or e2>0 set e3=1

Check that only valid codes are in

Variable name: e1_(1-12)

Question: Measured height in cm. Centimetres (with one decimal)

Universe: All persons less than 5 years

Valid codes:

Edits & imputations

if e1= SYSMIS and e3=1, set e1=999

Within each year-group 0-5 do range check: if e1 ne 999 and e1 >= 3 x stdv of MEAN then replace with MEDIAN within year-group and state. Flag. If flag, LIST and check TIFF manually for possible decimal error. If error, change manually and flag.

If b4_1 age >=5 set e1= SYSMIS

Variable name: e2_(1-12)

Question: Measured weight in kg. Kilograms (with one decimal)

Universe: All persons less than 5 years

Valid codes:

Edits & imputations

if e2= SYSMIS and e3=1, set e2=999

Within each year-group 0-5 do range check: if e2 ne 999 and e2 >= 3 x stdv of MEAN then replace with MEDIAN within year-group and state. Flag.

If flag, LIST and check TIFF manually for possible decimal error. If error, change manually and flag.

If b4_1 age >=5 set e2= SYSMIS

Run SPSS MVA or similar test on all variables for corrected module e1_e3 printout report and store temporary work file
--

4.3.5 Questionnaire module: F (ONLY FOR SSCCSE)

Syntax file name:	nbhs_08_check F
Input:	tmp_E_state#
4.4 Output:	TMPWORK[state#]_PERSON

PURPOSE

Clean the person file module F1-F3

Completeness and consistency checks. Imputation. Flagging of changes.

For persons 12-24 months old

Variable name: f1_(1-12)

Question: Has your 1 years old child [Name] ever been given measles vaccination injections or MMR (that is a "shot" in the arm given at the age of 9-12 months)?

Universe: All persons up to 12-24 months old

Valid codes: 1,2 and 9 (no response)

Edits & imputations

If month => 12 and month <25 then:

If f1 =9 (double entry from scanning) check TIFF manually and decide on 1 to 2. flag

LOOP until correct

if f1= 2 or f1= SYSMIS and f2=1 or f3=1 then set f1=1. Flag

if f1= SYSMIS then set f1=9

if b4_2 month ne 12-24 then set f1= SYSMIS

Check that only valid codes are in

Variable name: f2_(1-12)

Question: Is there a vaccination card to confirm this?

Universe: All persons up to 1 year old

Valid codes: 1,2 and 9 (no response)

Edits & imputations

If month => 12 and month <25 then:

If f2 =9 (double entry from scanning) check TIFF manually and decide on 1 to 2. flag

LOOP until correct

If f2=SYSMIS or f2=2 and f1=1 and f3=1 then set f2=1

if f2= SYSMIS and f1=1 and f3=SYSMIS set f2=9

if b4_2 month ne 12-24 then set f2= SYSMIS

Check that only valid codes are in

Variable name: f3_(1-12)

Question: Is there a vaccination card to confirm this?

Universe: All persons up to 1 year old

Valid codes: 1

Edits & imputations

If b4_2 month => 12 and month <25 then:

if f2=2 and f3=1 set f3=SYSMIS. Flag

if f2=1 and f3=SYSMIS set f3=9

if b4_2 month ne 12-24 then set f3= SYSMIS

Check that only valid codes are in

Run SPSS MVA or similar test on all variables for corrected module F printout report and store temporary work file
--

4.5 WORKFILE (c): Housing and household file (H-L + parts of N) file name: TMPWORK[state#]_HOUSE

PURPOSE

Structure and clean a household file (one record per household with corrected variables from module H-L. Completeness and consistency checks. Imputation. Flagging of changes.

EDITS/RESTRUCTURE:

- 1 Select all variables H-L with one record per household
- 2 Keep selected key variables from section A1-4 on each person record: #date, #time, num_1, tiff_(1 – 8), geo_code, household, a3_urbrur, a1_worker, all A2, all A3, a4_tot_form, a4_thisform, a4_month and a4_year
- 3 If no information filled in for module H-L, delete the whole record
- 4 Clean H-L and store file

4.5.1 Questionnaire module: H

Syntax file name:	nbhs_09_check F
Input:	tmp_2A_state#
Output:	tmp_H_state#

PURPOSE:

Clean the person file module H1-H10

Completeness and consistency checks. Imputation. Flagging of changes.

For all households

Variable name: h1

Question: What type of dwelling does this household live in? (mark only one)

Universe: All households

Valid codes: 1-11 and 99 (no response)

Edits & imputations

If h1 =9 (double entry from scanning) check TIFF manually and decide on 1 to 11. flag
LOOP until correct

If h1= SYSMIS, set h1=99

Check that only valid codes are in

Variable name: h2

Question: How many rooms does this household have total?

Universe: All households

Valid codes:

Edits & imputations

If h2= SYSMIS then set h2= 99

Within urban rural strata do range check: if h2 <>99 and h2 >= 3 x stdv of MEAN then replace with MEDIAN within each strata. Flag.

If flag, LIST and check TIFF manually for possible explanations. If error found, change manually and flag.

Variable name: h3

Question: How many rooms are used for sleeping indoors?

Universe: All households

Valid codes:

Edits & imputations

If h3= SYSMIS then set h3=99

If h3>h2 and h3 <> 99 and h2 <>99 and h2 not flagged, set h3=h2. Flag

If h3>h2 and h2 =flagged, LIST and manual check/correct h2 and h3. Flag

Variable name: h4

Question: What is the main tenure status of this dwelling?

Universe: All households

Valid codes: 1-4 and 9 (no response)

Edits & imputations

If h4 =9 (double entry from scanning) check TIFF manually and decide on 1 to 4. flag

LOOP until correct

If h4=SYSMIS, set h4=9

Check that only valid codes are in

Variable name: h5

Question: What is the main source of drinking water for this household?

Universe: All households

Valid codes: 1-13 and 99 (no response)

Edits & imputations

If h5 =99 (double entry from scanning) check TIFF manually and decide on 1 to 11. flag

LOOP until correct

If h5= SYSMIS, set h5=99

Check that only valid codes are in

Variable name: h6

Question: How long time does it usually take to walk (one way) to this main water source from dwelling (if source is in dwelling, enter 0)?

Universe: All households

Valid codes:

Edits & imputations

Within urban rural strata do range check: $h6 \geq 3 \times \text{stdv of MEAN}$ then replace with MEDIAN within each strata. LIST and Flag.

If h6= SYSMIS set h6=999

Variable name: h7

Question: What is the main source of lighting for this household?

Universe: All households

Valid codes: 1-11 and 99 (no response)

Edits & imputations

If h7 =99 (double entry from scanning) check TIFF manually and decide on 1 to 11. flag

LOOP until correct

If h7= SYSMIS, set h7=99

Check that only valid codes are in

Variable name: h8

Question: What is the main source of energy for cooking for this household?

Universe: All households

Valid codes: 1-9 and 99 (no response)

Edits & imputations

If h8=99 (double entry from scanning) check TIFF manually and decide on 1 to 11. flag

LOOP until correct

If h8= SYSMIS, set h8=99

Check that only valid codes are in

Variable name: h9

Question: What is the main type of toilet facility used for this household?

Universe: All households

Valid codes: 1-6 and 9 (no response)

Edits & imputations

If h9 =9 (double entry from scanning) check TIFF manually and decide on 1 to 6. flag

LOOP until correct

If h9= SYSMIS, set h9=9

Check that only valid codes are in

Variable name: h10

Question: What is the main method of solid waste disposal for this household?

Universe: All households

Valid codes: 1-6 and 9 (no response)

Edits & imputations

If h10 =9 (double entry from scanning) check TIFF manually and decide on 1 to 11. flag

LOOP until correct

If h10= SYSMIS, set h10=9

Check that only valid codes are in

Run SPSS MVA or similar test on all variables for corrected module H printout report and store temporary work file

4.5.2 Questionnaire module: I

Syntax file name: nbhs_10_check I

Input: tmp_H_state#

Output: tmp_I_state#

PURPOSE:

Clean the person file module I1-I3

Completeness and consistency checks. Imputation. Flagging of changes.

For all households

Variable name: i1

Question: What is the household's main source of livelihood?

Universe: All households

Valid codes: 1-9 and 99 (no response)

Edits & imputations

If i1 =99 (double entry from scanning) check TIFF manually and decide on 1 to 9. flag

LOOP until correct

If i1= SYSMIS, set i1=99

Variable name: i2_1 (1-5)

Question: Does any member of this household own any of the following transport items? (mark all that apply) If marked: How many items do you have and how much would you have to pay if you should by this item at the market?

Universe: All households

Valid codes: 1,2 and 9 (no response)

Edits & imputations

If i2_1 =9 (double entry from scanning) check TIFF manually and decide on 1 to 2. flag

LOOP until correct

if i2_1 =1 or SYSMIS and i2_2 and i2_3 =SYSMIS, set i2_1=2

if i2_1 = 2 or SYSMIS and i2_2 or i2_3 >0, set i2_1=1

Check that only valid codes are in

Variable name: i2_2

Question: Does any member of this household own any of the following transport items? (mark all that apply) If marked: How many items do you have and how much would you have to pay if you should by this item at the market?

Universe: All households

Valid codes:

Edits & imputations

if i2_2= SYSMIS and i2_1 = 1, set i2_2 = 999

Variable name: i2_3

Question: Does any member of this household own any of the following transport items? (mark all that apply) If marked: How many items do you have and how much would you have to pay if you should by this item at the market?

Universe: All households

Valid codes:

Edits & imputations

if i2_3= SYSMIS and i2_1=1, set i2_3 = 999

COMPUTE new variable “i2value” with 1 decimal = $i2_3/i2_2$ (if i2_2 and i2_3 > 0 and ne 999
Range check if i2value >= 3 x stdv of MEAN in State, replace with MEDIAN i2value. LIST. Flag.
Adjust i2_2 and i2_3. LIST. Flag

Variable name: i3_1 (1-5)

Question: Does any member of this household own any of the following? (mark all that apply) If marked: How many items do you have and how much would you have to pay if you should by this item at the market?

Universe: All households

Valid codes: 1,2 and 9 (no response)

Edits & imputations

If i3_1 =9 (double entry from scanning) check TIFF manually and decide on 1 to 2. flag
LOOP until correct

if i3_1 =1 or SYSMIS and i3_2 and i3_3 =SYSMIS, set i3_1=2

if i3_1 = 2 or SYSMIS and i3_2 or i3_3 >0, set i3_1=1

Check that only valid codes are in

Variable name: i3_2

Question: Does any member of this household own any of the following transport items? (mark all that apply) If marked: How many items do you have and how much would you have to pay if you should by this item at the market?

Universe: All households

Valid codes:

Edits & imputations

if i3_2= SYSMIS and i3_1 = 1, set i3_2 = 999

Variable name: i3_3

Question: Does any member of this household own any of the following transport items? (mark all that apply) If marked: How many items do you have and how much would you have to pay if you should by this item at the market?

Universe: All households

Valid codes:

Edits & imputations

if i3_3= SYSMIS and i3_1 = 1, set i3_3 = 999

COMPUTE new variable “i3value” with 1 decimal = $i3_3/i3_2$ (if $i3_2$ and $i3_3 > 0$ and ne 999
Range check if $i3value \geq 3 \times \text{stdv of MEAN in State}$, replace with MEDIAN $i3value$. LIST. Flag.
Adjust $i3_2$ and $i3_3$. LIST. Flag

Run SPSS MVA or similar test on all variables for corrected module I printout report and store temporary work file

4.5.3 Questionnaire module: J

Syntax file name: nbhs_11_check J
Input: tmp_I_state#
Output: tmp_J_state#

PURPOSE:

Clean the person file module J1-J3

Completeness and consistency checks. Imputation. Flagging of changes.

For all households

Variable name: j1

Question: Do members of this household have access to any health care facility when sick?

Universe: All households

Valid codes: 1,2 and 9 (non response)

Edits & imputations

If $j1=9$ (double entry from scanning) check TIFF manually and decide on 1 to 2. flag

LOOP until correct

if $j1=1$ or SYSMIS and $j2-j4 = \text{SYSMIS}$, set $j1=2$

if $j1=2$ or SYSMIS and $j2$ or $j3$ or $j4 > 0$, set $j1=1$

Check that only valid codes are in

Variable name: j2

Question: What type of health care facility do members of this household visit most often when sick?

Universe: All households

Valid codes: 1-7 and 9 (no response)

Edits & imputations

If $j2=9$ (double entry from scanning) check TIFF manually and decide on 1 to 7. flag

LOOP until correct

if $j2= \text{SYSMIS}$ and $j1=1$, set $j2=9$

if $j2>0$ and $j1=2$, set $j2= \text{SYSMIS}$

Check that only valid codes are in

Variable name: j3

Question: Does the facility provide free health care?

Universe: All households

Valid codes: 1,2 and 9 (no response)

Edits & imputations

If $j3=9$ (double entry from scanning) check TIFF manually and decide on 1 to 2. flag

LOOP until correct

if $j3= \text{SYSMIS}$ and $j1=1$, set $j3=9$

if $j3>0$ and $j1=2$, set $j3= \text{SYSMIS}$

Check that only valid codes are in

Variable name: j4a
Question: How long does it take members of this household to reach the above mentioned health facility?
Universe: All households
Valid codes: 1-5 and 9 (no response)

Edits & imputations

If j4 =9 (double entry from scanning) check TIFF manually and decide on 1 to 4. flag
LOOP until correct
if j4a= SYSMIS and j1=1 and j4b=SYSMIS, set j4a=9
Check that only valid codes are in

Variable name: j4b (*j4a and j4b is split due to error in questionnaire design*)
Question: How long does it take members of this household to reach the above mentioned health facility?
Universe: All households
Valid codes: 1 and 9 (no response)

Edits & imputations

if j4b= SYSMIS and j1=1 and j4a=SYSMIS, set j4=9
Check that only valid codes are in

Variable name: j5
Question: IF members of this household do not usually visit any of the facilities above mentioned (j2), what is the main type of medical help you seek when sick?
Universe: All households
Valid codes: 1-5 and 9 (no response)

Edits & imputations

If j5 =9 (double entry from scanning) check TIFF manually and decide on 1 to 5. flag
LOOP until correct
if j5= SYSMIS set j5=9
Check that only valid codes are in

Variable name: j6
Question: Does this household have any bed-/mosquito-nets?
Universe: All households
Valid codes: 1, 2 and 9 (no response)

Edits & imputations

If j6 =9 (double entry from scanning) check TIFF manually and decide on 1 to 2. flag
LOOP until correct
if j6=1 or SYSMIS, and j7-10= SYSMIS, set j6=2
if j6=2 or SYSMIS and j7 or j8 or j9 or j10 >0, set j6=1

Back check to i3 for consistency:

if j6=1 and i3_1_10=2 set i3_1_10=1 and i3_2_10= j7 and i3_3_10=999 and i3value_10=999
if j6=2 and i3_1_10=1 set j6=1 and j7=i3_3_10
Check that only valid codes are in

Variable name: j7
Question: How many bed/mosquito nets does this household have?
Universe: All households
Valid codes:

Edits & imputations

if j7=blank and j6=1 then j7=999
Range check: If j7> 2 per household member ("h_size" variable merged from workfile II), then
LIST and manual check. Flag corrections

Variable name: j8 (1-4)
Question: Who usually sleep under these bed/mosquito nets? (mark all that applies)
Universe: All households
Valid codes: 1 (multiple response) and 9 (no response)

Edits & imputations

if j8 (1-4) = SYSMIS and j6=1 then j8 (1-4)=9
Check that only valid codes are in

Variable name: j9
Question: Have these bed/mosquitonets been treated with chemicals less than 12 months ago?

Universe: All households
Valid codes: 1,2 and 9 (no response)

Edits & imputations

if j9 = SYSMIS and j6=1 then j9 = 9
Check that only valid codes are in

Variable name: j10
Question: Where did the household acquire these bed/mosquito nets?

Universe: All households
Valid codes: 1-3 and 9 (no response)

Edits & imputations

if j10 = SYSMIS and j6=1 then j10 = 9
Check that only valid codes are in

Run SPSS MVA or similar test on all variables for corrected module j printout report and store temporary work file

4.5.4 Questionnaire module: K

Syntax file name: nbhs_12_check K
Input: tmp_J_state#
Output: tmp_K_state#

PURPOSE:

Clean the person file module K1-K5

Completeness and consistency checks. Imputation. Flagging of changes.

For all households

Variable name: k1_1
Question: Has the household received cash or goods from food aid programs in the last 12 months?

Universe: All households
Valid codes: 1,2 and 9 (no response)

Edits & imputations

If k1_1=9 (double entry from scanning) check TIFF manually and decide on 1 to 2. flag LOOP until correct

if k1_1=1 or SYSMIS and k1_2 or k1_3 >0, then k1_1=1

if k1_1=2 or SYSMIS and k1_2 and k1_3 = SYSMIS, then set k1_1=2

Check that only valid codes are in

Variable name: k1_2

Question: Cash received?

Universe: All households

Valid codes:

Edits & imputations

if k1_2= SYSMIS and k1_1=1 and k1_3>0, then set k1_2=0

Range check if k1_2<> 0 and k1_2 >= 3 x stdv of MEAN in State, replace with MEDIAN k1_2.
LIST. Flag.

Variable name: k1_3

Question: Value of kind received?

Universe: All households

Valid codes:

Edits & imputations

if k1_3= SYSMIS and k1_1=1 and k1_2>0 then set k1_3=0

Range check if k1_3<> 0 and k1_3 >= 3 x stdv of MEAN in State, replace with MEDIAN k1_3.
LIST. Flag.

<Repeat this for k1 – k5>

Run SPSS MVA or similar test on all variables for corrected module k printout report and store temporary work file

4.5.5 Questionnaire module: L

Syntax file name: nbhs_13_check L

Input: tmp_K_state#

Output: tmp_L_state#

PURPOSE:

Clean the person file module L1-L10

Completeness and consistency checks. Imputation. Flagging of changes.

For all households

Variable name: l1

Question: Has any member of this household used cash in the past 7 days?

Universe: All households

Valid codes: 1, 2 and 9 (non response)

Edits & imputations

If l1=9 (double entry from scanning) check TIFF manually and decide on 1 to 2. flag

LOOP until correct

if l1= SYSMIS then set l1=9

Check that only valid codes are in

Variable name: l2

Question: Does any member of this household use cash for obtaining goods for the household that the household does not produce itself?

Universe: All households

Valid codes: 1, 2 and 9 (no response)

Edits & imputations

If l2=9 (double entry from scanning) check TIFF manually and decide on 1 to 2. flag

LOOP until correct

if l2= SYSMIS then set l2=9

Check that only valid codes are in

Variable name: I3

Question: If not using cash, how does the household obtain such goods?

Universe: All households

Valid codes: 1-3 and 9 (no response)

Edits & imputations

If I3=9 (double entry from scanning) check TIFF manually and decide on 1 to 3. flag

LOOP until correct

if I3= SYSMIS then set I3=9

Check that only valid codes are in

Variable name: I4

Question: In the last 12 months has any member of the household borrowed or obtained money that he/she had to repay?

Universe: All households

Valid codes: 1,2 and 9 (no response)

Edits & imputations

If I4=9 (double entry from scanning) check TIFF manually and decide on 1 to 2. flag

LOOP until correct

if I4=1 or I4= SYSMIS and I5 (1-5) and I6 and I7= SYSMIS, set I4=2

if I4=2 or I4=SYSMIS and I5 (1-5) or I6 or I7 >0, then set I4=1

Check that only valid codes are in

Variable name: I5_(1-5)

Question: Who did the household borrow from in the last 12 months?

Universe: All households

Valid codes: 1 (multiple response) or 9 (no response)

Edits & imputations

if I5 (1-5) = SYSMIS and I4=1 then set I5 (1-5) = 9

Check that only valid codes are in

Variable name: I6

Question: How much money did the household borrow the last 12 months?

Universe: All households

Valid codes:

Edits & imputations

if I6 = SYSMIS and I4=1 then set I6 = 999

Range check if I6 <> 0 and <> 999 and I6 >= 3 x stdv of MEAN in State, replace with MEDIAN I6. LIST. Flag.

Variable name: I7_(1-15)

Question: What is the main reason for borrowing or obtaining loan?

Universe: All households

Valid codes: 1 (multiple response) and 99 (no response)

Edits & imputations

if I7 (1-15) are all = SYSMIS and I4=1 then set I7(1-15) = 99

Check that only valid codes are in

Variable name: I8_(1-9)

Question: For those who did not borrow or obtain money the last 12 months (from I4): Why has no members of the household borrowed money in the last 12 months?

Universe: All households

Valid codes: 1 (multiple response) and 99 (no response)

Edits & imputations

if I8_(1-9) are all = SYSMIS and I4=2 then set I8(1-9) = 99

if I8_(1-9) one or more are =1 and I4=1, then set I8-(1-9) to SYSMIS

Check that only valid codes are in

Variable name: I9

Question: Does any member of the household have a bank account or a postal savings account?

Universe: All households

Valid codes: 1,2 and 9 (no response)

Edits & imputations

If I9=9 (double entry from scanning) check TIFF manually and decide on 1 to 2. flag

LOOP until correct

if I9 = SYSMIS then set I9= 9

Check that only valid codes are in

Variable name: I10__(1-10)_1

Question: Over the past five years, was the household severely affected by any of the following events? Yes/no

Universe: All households

Valid codes: 1, 2 and 9 (no response)

Edits & imputations

If I10_c01=9 (double entry from scanning) check TIFF manually and decide on 1 to 2. flag

LOOP until correct

if I10_c01=1 or I10_c01 =SYSMIS and c02=0 and c03=0 and c04=0 then set I10_c01= 2

if I10_c01=2 or I10_c01=SYSMIS and c02 or c03 Or c04 > 0 set I10_c01=1

Check that only valid codes are in

Variable name: I10__(1-10)_2

Question: Over the past five years, was the household severely affected by any of the following events? Rank the three most significant shocks.

Universe: All households

Valid codes: 1,2,3

Edits & imputations

1 if c01=1 and c02 > 3 the LIST and manual check TIFF file

Variable name: I10__(1-10)_3

Question: Over the past five years, was the household severely affected by any of the following events? Estimated value of the most significant shocks

Universe: All households

Valid codes:

Edits & imputations

if c03 =SYSMIS and c01=1 then set c03=999

Range check if I10_3 <> 0 and <> 999 and I10_3 >= 3 x stdv of MEAN in State, replace with MEDIAN I10_3. LIST. Flag.

Variable name: l10__(1-10)_3

Question: Over the past five years, was the household severely affected by any of the following events? What did you do in response to this shock to try to cope/regain yours former welfare?

Universe: All households

Valid codes: 1-25

Edits & imputations

if c04= SYSMIS and c01=1 then set c04=99

if c04 > 25 and < 99, LIST and check TIFF codes manually

Check that only valid codes are in

Run SPSS MVA or similar test on all variables for corrected module L printout report and store temporary work file

4.5.6 Questionnaire module: N1-3 + N6

Syntax file name: nbhs_14_check N1

Input: tmp_L_state#

Output: **TMPWORK[state#]_HOUSE**

PURPOSE

Clean the household agriculture file module N1-3 and N6

Completeness and consistency checks. Imputation. Flagging of changes.

For all households

Variable name: n1

Question: Does any member of the household currently own or use any agriculture land, forest and pasture land?

Universe: All households

Valid codes: 1, 2 and 9 (no response)

Edits & imputations

if n1= 9 (from scanning multiple mark edits) check TIFF and decide on which code 1,2 to enter

if n1=1 and n2 and n3 and n4 and n5= SYSMIS then set n1=2

if n1=2 or n1=SYSMIS and n2>0 or n3>=1 or sum n4>0 or sum n5>0 then set n1=1

if n1=SYSMIS and n2 and n3 and n4 and n5= SYSMIS then set n1=9

Check that only valid codes are in

Variable name: n2_(1-4)

Question: What is the tenure status of the land under cultivation?

Universe: All households with agriculture

Valid codes: 1 (multiple response) and 9 (no response)

Edits & imputations

if n2_ (1 to 4)= SYSMIS and n1=1 then set n2_(1-4) = 9

Check that only valid codes are in

Variable name: n3_(1-5)

Question: How much did the household spend on the following agricultural inputs last 12 months?

Universe: All households with agriculture

Valid codes:

Edits & imputations

if n3_ (1 to 5)= SYSMIS and n1=1 then set n3_(1-5) = 0

Range check if n3_(1-5) <> 0 and and N3_(1-5) >= 3 x stdv of MEAN in State, replace with MEDIAN N3_(1-5). LIST. Flag.

Variable name: n6_1

Question: Does any member of the household currently own any livestock or poultry?

Universe: All households

Valid codes: 1, 2 and 9 (no response)

Edits & imputations

If n6_1=9 (double entry from scanning) check TIFF manually and decide on 1 to 2. flag
LOOP until correct

if n6_1=1 and n6_2_(1-8)= SYSMIS then set n6_1=2

if n6_1=2 or n6_1= SYSMIS and n6_2_(1-8)>0 0 then set n6_1=1

if n6_1=SYSMIS and n6_2_(1-8)= SYSMIS then set n6_1=9

Check that only valid code is in

Variable name: n6_2_1

Question: Does any member of this household own any of the following animals? (mark all that apply) Yes No

Universe: All households with livestock or poultry

Valid codes:

Edits & imputations

If n6_2_1=9 (double entry from scanning) check TIFF manually and decide on 1 to 2. flag
LOOP until correct

if n6_2_1(1-8)=1 or SYSMIS and n6_2_2(1-8)= SYSMIS and n6_2_3_(1-8)= SYSMIS then set
n62=2

if n6_2_1(1-8)=2 or SYSMIS and n6_2_2(1-8)>0 or n6_2_3(1-8)>0 then set n62=1

Variable name: n6_2_2

Question: Number of animals

Universe: All households

Valid codes:

Edits & imputations

if n6_2_2 = SYSMIS and n6_2_1=1 then set n6_2_2=999

Range check if n6_2_2 <> 0 and and n6_2_2 >= 3 x stdv of MEAN in State, replace with MEDIAN
n6_2_2. LIST. Flag.

Variable name: n6_2_3

Question: If you sold one of the animals today, how much money could you get for it?

Universe: All households with animals

Valid codes:

Edits & imputations

if n6_2_3= SYSMIS and n6_2_1=1 then set n6_2_3=999

Range check if n6_2_3 <> 0 and and n6_2_3 >= 3 x stdv of MEAN in State, replace with MEDIAN
n6_2_3. LIST. Flag.

Run SPSS MVA or similar test on all variables for corrected module n printout report and store temporary work file
--

4.6 WORKFILE (d): commodities purchased last 7 or 30 days. Commodity file (M1-3)

File name: TMPWORK[state#]_FOOD

PURPOSE

Create dummy variables for M1-M3 to enable for transpose. Transpose and clean a commodity file (one record per commodity with corrected variables from module M1-M3. Completeness and consistency checks. Imputation. Flagging of changes.

EDITS/RESTRUCTURE:

- 1 Select all variables M1-M3 with one record per commodity
- 2 Keep selected key variables from section A1-4 on each person record: #date, #time, num_1, tiff_(1 – 8), a1_state, geo_code, household, a3_urbrur, a1_worker, all A2, all A3, a4_tot_form, a4_thisform, a4_month and a4_year and “household_size” from corrected section B (work file II)
- 3 If no information filled in for module M1-M3, delete the whole record
- 4 Create new variable “item_code”
- 5 Clean M1-M3 and store workfile IV with derived/converted commodity codes, commodity purchased and/or consumed in kg, total cash for purchase, place of purchase and estimated price, flags + selected key variables A, B and M1

4.6.1 Questionnaire module: M1

Syntax file name: nbhs_15_check M1

Input: tmp_2A_state#

Output: tmp_M1_state#

PURPOSE:

Clean the commodity file M1 and transpose M1-3

Completeness and consistency checks. Imputation. Flagging of changes.

For all households

Variable name: M1_1

Question: Does this household have more than 12 members?

Universe: All households

Valid codes: 1,2 and 9 (no response)

Edits & imputations

Create dummivariabls for M3 (for all the shaded fields in section m3) before transpose TRANSPOSE M2-3 and keep selected and corrected variables from A, B (household_size) and M1_n each record. Create new variable “item_code”

If M1_1=9 (double entry from scanning) check TIFF manually and decide on 1 to 2. flag LOOP until correct

if M1_1=1 or SYSMIS and “household_size” <=12, then set M1_1=2 and LIST and check if M1_1=2 or SYSMIS and “household_size” >12, then set M1_1=1

Variable name: M1_2

Question: Is this the last form used for the first visit interview of the household?

Universe: All households

Valid codes: 1, 2 and 9 (no response)

Edits & imputations

If M1_2=9 (double entry from scanning) check TIFF manually and decide on 1 to 2. flag LOOP until correct

Variable name: M1_3

Question: How many people ate together at a daily basis in the household during the last 7 days?

Universe: All households

Valid codes:

Edits & imputations

if m1_3 = SYSMIS, set m1_3= "household_size", flag

if m1_3 > "household_size", LIST and check TIFF file

Run SPSS MVA or similar test on all variables for corrected module m1 printout report and store temporary work file

4.6.2 Questionnaire module: M2-M3

Syntax file name: nbhs_16_check M2

Input: tmp_m1_state#

Output: **TMPWORK[state#]_FOOD**

PURPOSE

Clean the commodity file M2-3 perform conversion to kg

Completeness and consistency checks. Imputation. Flagging of changes.

Variable name: M2_c02

Question: Mark yes or no for all items.

Universe: All items

Valid codes: 1, 2

Edits & imputations

If M2_c02=9 (double entry from scanning) check TIFF manually and decide on 1 to 2. flag LOOP until correct

if all c03 to c15 = SYSMIS then delete record

if c02=2 or c02=SYSMIS and (c03 or c04 or c05>0) or (c07 and c08>0) or (c09 and c10>0) or (c11and c12>0) or (c13 and c14>0) then set c02=1. Flag

Variable name: M2_c03, c04 / c07, c08 / c09, c10 / c11, c12 / c13, c14

Question: Quantity and unit of quantity

Universe: All items

Valid codes:

Edits & imputations & conversion

Step 1 edits:

if c03/07/09/11/13 = 50 and corresponding c04/08/10/12/14=17, set c03/07/09/11/13=1. flag (this is ruling out a common misunderstanding of 50 sacks vs unit sack 50kg. flag

if c06=SYSMIS and c03 or c04 or c05 >0 then set c06=9

Based on the prepared conversion table (nhbs_lookup2), COMPUTE new variables:

- purchase_kg,
- consume_purchase_kg,
- consume_stock_kg,
- consume_production_kg,
- consume_gifts_kg

if consume_purchase_kg (c07&c08) > purchase_kg (c03&c04) then set consume_purchase_kg= purchase_kg

if c03-05, c07-14 not a "zero or SYSMIS division" calculate the derived (typeB) variables:

COMPUTE new variables (B):

- total_consume_kg=(consume_purchase_kg+consume_stock_kg+consume_production_kg + consume_gifts_kg)
- purchased_kg_per_person,
- consumed_kg_per_person,
- stock_kg_per_person,
- production_kg_per_person,
- gifts_kg_person

Step 2 edits & Checks & Calculations

Calculate and LIST item characteristics for all derived (B) variables (N, max, min, mean, median, stdv, quartile)

For urban&rural strata and commodities with 30 observations or more on variable total_consume_kg", run outlier range check: if "new variable" ≥ 3 stdv from MEDIAN replace with MEDIAN. LIST. Flag

For commodities flagged for median, recalculate related variables to achieve consistency. flag

For "new variables" with less than 30 observations, aggregate within commodity 3digit level, flag for aggregate and run the same outlier test on the aggregated variable. LIST. Flag

Variable name: M3

Question: Purchase and consumption last 30 days?

Universe: All households

Edits&imputations:

Special case of M2.

Run SPSS MVA or similar test on all variables for corrected module m2-3 printout report and store temporary work file

4.7 WORKFILE (e): commodities purchased last 30 days or last year. Commodity file (M4-5) TMPWORK[state#]_GOODS

PURPOSE

Transpose and clean a commodity file (one record per commodity with corrected variables from module M4-5. Completeness and consistency checks. Imputation. Flagging of changes.

EDITS/RESTRUCTURE:

- 1 Select all variables M4-5 and transpose to one record per commodity
- 2 Keep selected key variables from section A1-4 on each person record: #date, #time, num_1, tiff_(1 – 8), a1_state, geo_code, household, a3_urbrur, a1_worker, all A2, all A3, a4_tot_form, a4_thisform, a4_month and a4_year and "household_size" from corrected section B (work file II)
- 3 If no information filled in for module M4-5, delete the whole record
- 4 Create new variable "item_code" and "periode"
- 5 Clean M4-M5 and store workfile with commodity codes, commodity purchased, total cash for purchase, flags + selected key variables A, B and M1

4.7.1 Questionnaire module: M4_M5

Syntax file name: nbhs_17_check M3

Input: tmp_2A_state#

Output: **TMPWORK[state#]_GOODS**

PURPOSE

Transpose and clean the commodity file M4-5 perform

Completeness and consistency checks. Imputation. Flagging of changes.

New variables:

- 1 Create new variable "m45_periode" and "m45_item"
- 2 For all commodities under m4 set m45_periode=1 (month)
- 3 For all commodities under m5 set m45_periode=2 (year)
- 4 For all commodities calculate "m45_item" (commodity code)

Transpose M4-5 to one line per item

Keep selected variables from A,B, flags + new variables created under m45 and all screening questions on each record

Variable name: m4_skip_(562-121)

Question: Mark yes and no

Universe: m4 Commodity group (562 – 121)

Valid codes: 1,2

Edits & imputations

if m4_skip_(562-121)=1 or m4_skip_(562-121)= SYSMIS and m4_c03_(562_121)= SYSMIS, delete skip and all records in skip group

if m4_skip_(562-121)= 2 or m4_skip_(562-121)= SYSMIS and m4_c03_(562_121)>0, set m4_skip_(562-121)=1

Variable name: m4_c02_(562-121) and m4_c03:

Question: Mark yes and no

Universe: m4 Commodity group (562 – 121)

Valid codes: 1,2

Edits & imputations

1 if m4_c02_(562-121) = 1 and m4_c03_(562_121)= SYSMIS, set m4_c02=2

2 if m4_c02_(562-121) = 2 or SYSMIS and m4_c03_(562_121)>0, set m4_c02_(562-121)=1

For urban&rural strata and commodities with 30 observations or more on variable

c03 >0, run outlier range check: if "c03" >=3 stdv from MEDIAN replace with

MEDIAN. LIST. Flag

For "c03 variables >0 " with less than 30 observations, aggregate within commodity 3digit level, run the same outlier test on the aggregated variable. LIST. Flag

<Repeat same procedures for m5>

Run SPSS MVA or similar test on all variables for corrected module m4-m5 printout report and store temporary work file

4.8 WORKFILE (f) CROP PLOTS File name: TMPWORK[state#]_PLOTS

4.8.1 Questionnaire module: N4

Syntax file name: nbhs_18_check n4

Input: tmp_2A_state#

Output: **TMPWORK[state#]_PLOTS**

PURPOSE:

Select file and add ID from A, household_members and store ID + N variables temporary

Transpose n4 to one line per plot and keep ID variables and n1-3 per plot line. Calculate new variable "tot_area_ha". Check consistency with agric screening question in the household file

Variable names: n4_2, n4_4, n4_5, n4_8, n4_9, n4_10

Question:

Universe: All households with yes in screening variable n1 in the household module

Valid codes:

Edits & imputations

- 1 if any of the marker variables are coded 9 (more than one mark in a single response question) then LIST and check manually with the TIFF file, select only one mark and type in that value replacing the code 9.

Variable name: n4_(1-6)_c02

Question: Plot area unit code

Universe: All households with n1=1

Valid codes: 1,2,3 and 99 (no response)

Edits & imputations

if n4_(n)_c02= SYSMIS and n4_(n)_c03= SYSMIS and n4_(n)_c04= SYSMIS and n4_(n)_c05= SYSMIS then DELETE record

if n4_c02 >0 or n4_c03>0 or n4_c04>0 or n4_c05>0 then set all SYSMIS n4_(n)_c02-c11)=99

if n4_c02=1,2 or 3 and n4_c03>0 convert and calculate new variable n4_hectar_(1-6) with 3 decimals

if n4_hectar_(1-6) > 0 and n4_11_(1-6)>0, calculate new variable n4_sdgperhectar_(1-6)

Range check: if n4_sdgperhectar >= 3x stdv from MEDIAN and impute median value. LIST and flag.

For all LISTED, correct manually either n4_hectar_(1-6) or n4_11_(1-6) to match n4_sdgperhectare imputation. LIST and flag

Variable name: n4_(1-6)_c06 and c07

Question: crop code for first and second season respectively

Universe: All households with n1=1

Valid codes: 1-22 or 99

Edits & imputations

If n4_(1-6)_c06 > 22 and ne 99, then LIST and check TIFF manually

<same procedure for c07>

Check that only valid codes are in

Variable name: tot_area_ha

Question: new variable

Universe: All households with n1=1

Valid codes:

Edits & imputations

COMPUTE "tot_area_ha" = sum n4_hectar per household – hectare with 3 decimals

Run SPSS MVA or similar test on all variables for corrected module n4 printout report and store temporary work file

4.9 WORKFILE (g) CROP HARVEST File name: TMPWORK[state#]_HARVEST

4.9.1 Questionnaire module: N5

Syntax file name: nbhs_19_check n5
Input: tmp_2A_state#
Output: **TMPWORK[state#]_HARVEST**

PURPOSE:

Select file and add ID from A, household_members and store ID + N variables temporary
Transpose n5 to one line per croptype and keep ID variables and "tot_area_ha" per plot line.
Compute new variable n5_c01=line number . Check consistency with agric screening question "n1"
in the household file

Variable names: n5_c01

Question:

Universe: All households with yes in screening variable n1 in the household module

Valid codes:

Edits & imputations

COMPUTE n5_c01=line number

Variable names: n5_c03

Question: Have you harvested any [crop type] during the past 12 months?

Universe: All households with yes in screening variable n1 in the household module

Valid codes: 1,2

Edits & imputations

If N5_c03=2 or SYSMIS and c04-c08= SYSMIS then delete the record

Variable names: n5_c04 and n5_c05

Question: How much [crop type] did you harvest in the past 12 months?

Universe: All households with yes in screening variable n1 in the household module

Valid codes:

Edits & imputations

Convert c05 quantity unit code to SI unit kg based on lookup table "nhbs_lookup2"

COMPUTE new variable "kg_harvest" from c04 and converted c05

Range check: Per commodity type and if >= 30 observations: if kg_harvest>=3 x stdv from MEDIAN
replace with MEDIAN

If < 30 observations aggregate on similar crop types and run the same range check LIST and Flag

Aggregates:

- Cereales (1-5)
- Root&Tubers (6-9)
- Beans&peas (10-14)
- Vegetables (15-18)
- Other (19)

Run SPSS MVA or similar test on all variables for corrected module n5 printout report and store temporary work file

4.10 WORKFILE (h) HOUSEHOLD INCOME File name: TMPWORK[state#]_INCOME

4.10.1 Questionnaire module: O

Syntax file name:	nbhs_20_check o
Input:	tmp_2A_state#
Output:	TMPWORK[state#]_INCOME

PURPOSE:

Select file and add ID from A, household_members and store ID + N variables temporary
One line per household. Check consistency and outliers.

Variable names: o1_1_(1-8) and o1_2_(1-8)

Question: Household agric income during last month (by source 1-8) and by month(1) or year(2). (SDG)

Universe: All households (CBS)

Valid codes:**Edits & imputations**

COMPUTE new variable agric_month = SUM 1-8 (1) and agric_year=SUM 1-8(2)

Range check: For urban&rural strata if agric_month (agric_year) >=3 x stdv from MEDIAN
replace with MEDIAN. LIST and flag Look for 000 errors

Check consistency with N1:

If n1=1 and agric_month = SYSMIS and agric_year=SYSMIS LIST

If n1=2 and agric_month > 0 and agric_year>0 LIST

If inconsistency, change N1-3 and/or n6.1-2

Variable names: o2_1_(1-10) and o2_2_(1-10)

Question: Household non-agric income during last month (by source 1-10) and by month(1) or year(2). (SDG)

Universe: All households (CBS)

Valid codes:**Edits & imputations**

COMPUTE new variable nonagric_month = SUM 1-10 (1) and nonagric_year=SUM 1-10(2)

Range check: For urban&rural strata if nonagric_month (nonagric_year) >=3 x stdv from MEDIAN
replace with MEDIAN. LIST and flag Look for 000 errors

Run SPSS MVA or similar test on all variables for corrected module o printout report and store temporary work file
--