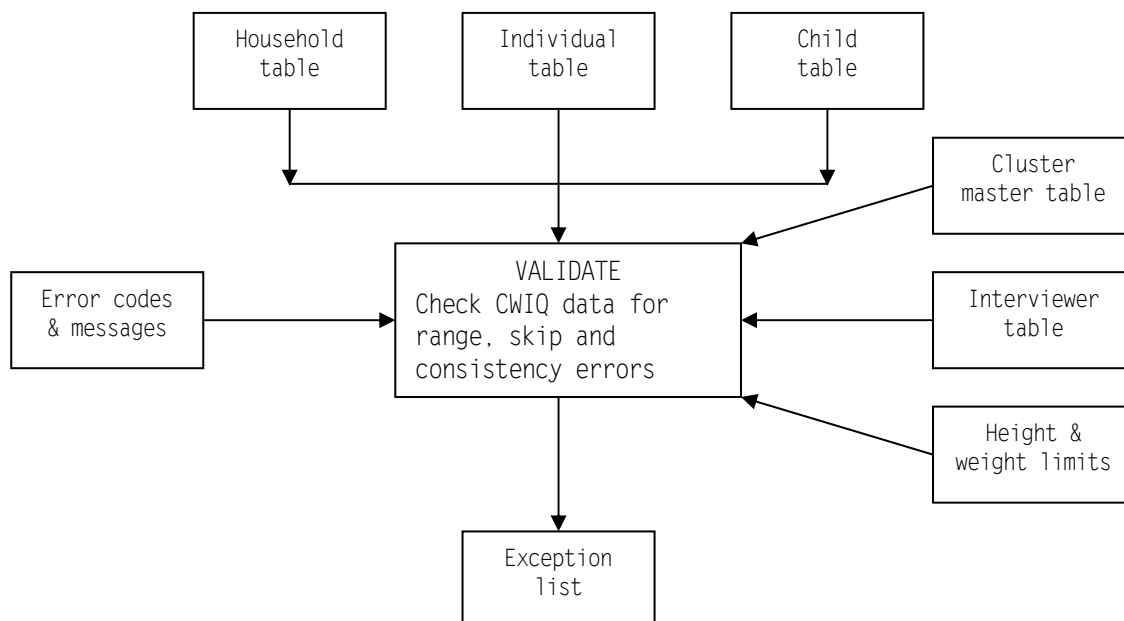


Program: Validate - Check CWIQ data consistency

1. Schematic



2. Input

A. CWIQ questionnaire database.

The CWIQ questionnaire database consists of five tables: the household table, the agriculture module table, the individual table, the child table for children under 5 and the woman table for women age 15 to 49. For each household surveyed there is one record in the household table (HHData) and the Subjective Poverty module (SPData), one record for each household member in the individual table (INData) and one record for each child under 5 in the child table (CHData).

B. Error codes and messages.

This table contains an entry for every exception detected by the program. Each entry is identified by the error code assigned to the exception; it also contains the text of the message to be printed in the exception list.

C. Cluster master table.

This table is used to validate the cluster number (A1). It contains an entry for all clusters in the sample.

D. Interviewer table.

This table is used to validate the interviewer number (A3). It contains an entry for all interviewers participating in the survey.

E. Height and weight limits.

This table is used to check the height and weight recorded for children under 5. It contains minimum and maximum height and minimum and maximum weight by age and sex of the child. The range of values is used to identify and correct, if necessary, extreme values of height and weight.

3. Output

A. Exception list.

The exception list shows all exceptions detected by the program. For each questionnaire with an exception, the program lists the questionnaire identification and the relevant data. Messages showing the exception code and the reason for the exception are printed after the data. The exception list is compared to the questionnaire to determine the cause of the exception and the way to correct it. A sample exception list appears in Appendix B.

4. Processing

Validation is done by cluster except at the end of the survey. The operator selects the cluster to process.

Data for each household in the cluster are extracted from the three tables and stored in memory. The data in memory are checked for each condition listed in the validation specifications below. When an exception is detected an error code is stored in a list associated with the type of data (household, individual and child). For questionnaires with detected errors, the relevant data, error codes and error messages are written to the exception list. A data and error summary are printed after all households in the cluster have been processed.

Validation specifications

Range checks

Range checks for each question are specified in the questionnaire flowchart.

Section A - Interview information

A1 – cluster number

The cluster number is checked against the cluster master table.

A3 – number of questionnaires

The number of questionnaires is a function of the number of household members. It should be equal to the number of household members + 9 divided by 10 (integer division).

A4 – interviewer number

The interviewer number is checked against the interviewer table.

A5 - interview date

The range is defined by the survey calendar, i.e. by the field work start and end dates.

A7 - respondent number

The range is 1-n, where n is the number of household members.
Households with a respondent under a certain age (say 15) should be identified for verification.

A6, A9 - interview start and end times

The range is country specific. A9 (end time) must be after A5 (start time).
Extreme values of interview duration [A9 - A5] (say over 90 minutes) should be signalled.

A8 – result of interview

Complete questionnaires must have household members and data in sections F, G, H and K.
Incomplete questionnaires must have no household members, no children under 5 and no data in sections F, G, H and P.

Section B - Household member general information

Member numbers must be consecutive starting with 1.

The head of household should be listed as member number 1.

B1, B2, B3 and B4 are asked of all members.

B3 (relation to head of household) and member number

Member 1 and only member 1 should be the head of household

B3 (relation to head of household) and B4 (age)

? hh level ?

B3 = 1 (head)	age > minimum (say 15)
B3 = 3 (child of head)	age < age of head - minimum (say 12)
B3 = 4 (parent of head)	age > age of head + minimum (say 12)

? should we check age of head against age of spouse ?

B4 (age) and B5 (marital status)

B4 < 10 implies that B5 is not asked
B4 >= 10 implies that B5 is asked

B3 (relation to head of household) and B5 (marital status)

B3 = 2 (spouse of head) implies that B5 = 2 or 3 (married)

B4 (age) and B6 (is father alive?)

B4 > 18 implies that B6, B7, B8, B9 are not asked
B4 <= 18 implies that B6, B8 are asked

B6 (is father alive) and B7 (is father living in the household)

B6 = 1 (yes) implies that B7 is asked
B6 <> 1 implies that B7 is not asked

B8 (is mother alive) and B9 (is mother living in the household)

B8 = 1 (yes) implies that B9 is asked
B8 <> 1 implies that B9 is not asked

At household level (i.e. for all members):

The number of household members must be the same as the household level variable, HHSIZE.

There must be one and only one head of household..

The presence of a spouse (B3 = 2) implies that the head of household is married (B5 = 2, 3).

The presence of two or more spouses (B3 = 2) implies that the head of household and the spouses are polygamous (B5 = 3)

The spouse (B3 = 2) of the head of household should have the same marital status (B5) as the head.

The spouse (B3 = 2) of the head of household should have the opposite sex (B1) of the head.

Section C - Household member education

B4 (age) and C1 (literacy)

B4 (age) < 15 implies that C1 is not asked (not applicable)

B4 (age) >= 15 implies that C1 is asked

C2 is asked of all members

C2 (school attendance)

C2 = 2 (no) implies that C3-C9 are not asked

C2 = 1 (yes) implies that C3-C5 are asked

B4 (age) and C3 (highest grade completed)

C3 = 01 (preschool)

B4 = 3-8

C3 = 11-16 (primary)

B4 = 5-15

C3 = 21-26 (secondary)

B4 = 10-20

C3 = 31, 41-43 (post-secondary)

B4 = 15+

C5 (currently in school)

C5 = 2 (no) implies that C6-C8 are not asked and C9 is asked.

C5 = 1 (yes) implies that C6-C8 are asked and C9 is not asked.

B4 (age) and C6 (current level of school)

C6 = 01 (preschool)

B4 = 3-8

C6 = 11-16 (primary)

B4 = 5-15

C6 = 21-26 (secondary)

B4 = 10-20

C6 = 31, 41-43 (post-secondary)

B4 = 15+

C6 (current grade) and C3 (highest grade completed)

C6 (current) >= C3 (highest completed)

C8 (problems with school)

C8a = 1 (no problem) implies that C8b-C8j are not marked.

C8a <> 1 implies that at least one of C8b-C8j is marked.

C5 (currently in school) and C9 (reason not in school)

C5 = 1 (currently in school) implies that C9 is not asked.

C5 = 2 (not currently in school) implies that C9 is asked.

If C9 (reason not currently in school) is asked

C9a = 1 (completed school) implies that C9b-C9m are not marked.
C9a <> 1 implies that at least one of C9b-C9m is marked.

C2 (never attended school), (B4) age and C10 (reason not started school)

C2 = 1 (attended school) implies that C10 is not asked.
C2 = 2 (never attended school) and age > 18 implies that C10 is not asked.
C2 = 2 (never attended school) and age <= 18 implies that C10 is asked.

Section D - Household member health

B1 (sex) and B4 (age) and D1

B1 <> 2 (not female) or B4 < 12 or B4 > 49 implies that D1 is not asked.
B1 = 2 (female) and B4 in 12-49 implies that D1 is asked.

D1 (birth in last year) and D2

D1 <> 1 (no birth in last 12 months) implies that D2 is not asked.
D1 = 1 (birth in last 12 months) implies that D2 is asked.

D3 (sick, injured in last 4 weeks)

D3 = 2 (no) implies that D4 is not asked.
D3 = 1 (yes) implies that D4 is asked.

D5 (consulted in last 4 weeks)

D5 = 2 (no) implies that D6-D8 are not asked.
D5 = 1 (yes) implies that D6-D8 are asked

D8 (problems with consultation)

D8a = 1 (no problem) implies that D8b-D8i are not marked.
D8a <> 1 implies that at least one of D8b-D8i is marked.

D5 (consulted in last 4 weeks) and D9 (reason for no consultation)

D5 = 1 (yes) implies that D9 is not asked.
D5 = 2 (no) implies that D9 is asked.

D9 (reason for no consultation)

D9a = 1 (no need) implies that D9b-D9e are not marked.
D9a <> 1 implies that at least one of D9b-D9e is marked.

Section E - household member employment

B4 (age) and Section E

B4 (age) < 5 implies that Section E (E1-E11) is not used.

B4 (age) ≥ 5 implies that E1 is asked.

E1 (paid work in the last 7 days)

E1 = 1 (yes) implies that E2-E5 are not asked
and that E6-E11 are asked.

E1 = 2 (no) implies that E2 is asked.

E2 (paid or unpaid work in last week for at least 1 hour)

E2 = 1 (yes) implies that E3-E5 are not asked
and that E6-E11 are asked

E2 = 2 (no) implies that E3 is asked
and that E6-E11 are not asked

E3 (looking for work and ready for work in the last 7 days)

E3 = 1 (yes) implies that E4 is asked
and that E5 is not asked

E3 = 2 (no) implies that E4 is not asked
and that E5 is asked

E5 (main reason for not working or looking for work in last 7 days) and B4 (age)

E5 = 4 (student) implies that age < minimum (say 25) and is still in school (C5 = yes)

E5 = 7 (too young) implies that age < minimum (say 15)

E5 = 6 (too old) implies that age > maximum (say 65)

Section F - Household Assets

F4 (land area compared to 1 year ago) and F3 (area currently owned)

F3 = 0 implies that F4 = 3 (more) is not a valid response.

F5 (use other land) and F6 (area of other land used)

F5 = 1 (no) implies that F6 is not asked.

F5 <> 1 implies that F6 is asked.

F7 (other land area compared to 1 year ago) and F3 (other land area currently used)

F6 = 0 implies that F7 = 3 (more) is not a valid response.

F8a-g (agricultural inputs)

F8a (none) = 1 (yes) implies that F8b to F8g are not marked

F8a (none) <> 1 (yes) implies that at least one of F8b to F8g is marked

F8a (use of agriculture inputs) and F9 (main source of agricultural inputs)

F8a (none) = 1 (yes) implies that F9 is not asked

F8a (none) <> 1 (yes) implies that F9 is asked

F10an (number of livestock owned) and F10bn (owned now compared to 12 months ago) n=1..7

F10an = 0 implies that F10bn can not be 1 (more now)

Section G - Household Amenities

G11a-l (anti-malaria measures)

G11a = 1 (no measures) implies that G11b-l are not marked.

G11a <> 1 (measures taken) implies that at least one of G11b-l is marked.

Section H – Displacement/Food Aid

H1 (household displaced since 1990)

H1 = 1 (yes) implies that H2 is asked
H1 = 2 (no) implies that H2-H4 are not asked.

H2 (returned to place of origin)

H2 = 1 (yes) implies that H3 is asked and H4 is not asked.
H2 = 2 (no) implies that H3 is not asked and H4 is asked

H4a, H4b, H4c (reasons for not returning to place of origin)

H4a (first choice) must be marked

H4b not marked implies that H4c is not marked

H5 (any member received food aid in the last 12 months)

H1 = 1 (yes) implies that H6 is asked
H1 = 2 (no) implies that H6 is not asked.

H7a1 (monetary assistance received)

H7a1 = 1 (yes) implies that H7a2 is asked
H7a1 = 2 (no) implies that H7a2 is not asked.

H7b1 (education assistance received)

H7b1 = 1 (yes) implies that H7b2 is asked
H7b1 = 2 (no) implies that H7b2 is not asked.

H7c1 (medical assistance received)

H7c1 = 1 (yes) implies that H7c2 is asked
H7c1 = 2 (no) implies that H7c2 is not asked.

H7d1 (construction/building assistance received)

H7d1 = 1 (yes) implies that H7d2 is asked
H7d1 = 2 (no) implies that H7d2 is not asked.

H7e1 (water/sanitation assistance received)

H7e1 = 1 (yes) implies that H7e2 is asked
H7e1 = 2 (no) implies that H7e2 is not asked.

H7f1 (agricultural assistance received)

H7f1 = 1 (yes) implies that H7f2 is asked
H7f1 = 2 (no) implies that H7f2 is not asked.

If H8 (household participation in community support systems)

H8a = 1 (none) implies that H8b-H8h are not marked.
H8a <> 1 implies that at least one of H8b-H8h is marked.

Section I - Child Roster

B4 (age) and Section I

B4 (age) ≥ 5 implies that Section I (I1-I14) is not used.

B4 (age) < 5 implies that I1-I14 have valid answers.

I1b (Mother's member number)

The member indicated must be female (B1 = 2) and age 13 or above (B4 ≥ 13).

I2 (Date of birth of child)

I2 must be have at least a valid year and month, before the interview date (A4).

Age in years calculated from A4 and I2 should be within 1 year of the age (B4) given in Section B.

I6d (DPT2) and I6c (DPT1)

I6d = 1 or 2 (yes) implies that I6c = 1 or 2 (yes)

I6e (DPT3) and I6d (DPT2)

I6e = 1 or 2 (yes) implies that I6d = 1 or 2 (yes)

I6h (Polio2) and I6g (Polio1)

I6h = 1 or 2 (yes) implies that I6g = 1 or 2 (yes)

I6i (Polio3) and I6h (Polio2)

I6i = 1 or 2 (yes) implies that I6h = 1 or 2 (yes)