CHRIS Study

Interview – Personal data

Version 1.1

24th April 2024

Authors: LB, MG

1. Introduction

This module stores information related to the education level, the number of siblings and offspring, and place of origin, of the participants, that were collected at the interview.

Participants book a morning appointment at the CHRIS study center, ranging from 7.45 to 8.45 a.m. Each study participant is assigned a workflow at the reception. If there are ten study participants (maximum capacity), there are ten different workflows, marked with the letters from "A" to "K". The current workflow is as follows: A-B-C-D-E-F-G-H-I-K. All the workflows can be found in the documentation of CHRIS Baseline/General information/Administrative data, in the file named "Workflows at baseline assessment". The interview occurs always after the spiralography and the blood drawing, for most as the last session, after the ECG assessment and the self-administered questionnaire (workflows B, C, E, F, H, I, L). For the remainder, the interview occurs after breakfast and just before the self-administered questionnaire (workflows A and G) or in between the blood drawing and the anthropometry (workflow D).

The interview full text and its corresponding answer lists are available at CHRIS Baseline/Interview.

2. History version changes

Version 1 of this interview module was in use between August 24th,2011 and November 2nd, 2012, whereas Version 2 had been in use since November 5th, 2012.

Between the different versions, the following changes have occurred:

Version 1 to Version 2:

variables dropped: x0pe02, x0pe03, x0pe07, x0pe10

variables added: x0pe02a, x0pe02b, x0pe02c, x0pe02d, x0pe11a, x0pe11b

question rephrased: x0pe04 (DE, IT), x0pe08 - question and response options were rephrased (DE, IT), x0pe09 (DE, IT).

other: x0pe06 - response options 2 and 3 were collapsed.

The question order was not changed and the filtering criteria remained the same between the two versions. As reported, the rephrasing of some questions and answers has occurred in both language versions of the interview.

Furthermore, the cleaning process resulted in the following variables:

variables dropped: x0pe05a x0pe05b x0pe05c

variables added: x0pe05d x0pe05e x0pe11

3. Data cleaning

1. The main CHRIS dataset was loaded.
2. The variable of number of years of education, x0pe09, was corrected according to the content of the nurse's notes related to education, x0pen3. If the participant reported had reported eight years of education, i.e. primary and lower secondary school completed, and in the notes additional upper secondary school were reported, these additional years were added.
3. Similarly, the variable on the completed level of education, x0pe08, was corrected according to the content of x0pen3. If the participant reported nine to twelve years of education and the notes mentioned attending some type of vocational secondary school, the completed level of education was adjusted to "Vocational upper secondary school". If instead the participant reported nine to twelve years of education and the notes mentioned attending vocational secondary school without completing it, the completed level of education was adjusted to "Lower secondary school".
4. The municipality of origin collected in the interview, x0pe05a, was standardized so that each location was uniquely identified by a single name (e.g. German and Italian names of South Tyrol locations, or different abbreviations of the complete names). The municipality was then encoded into province and country of origin, x0pe05b and x0pe05c, respectively. They were further classified into macrocategories (municipalities within Venosta, other South Tyrol, Italy, each European Country, continents otherwise) and finer categories (municipalities within South Tyrol, each province of Italy, each Country).
5. The number of siblings, x0pe02, when missing, was computed summing the number of brothers, x0pe02a, and the number of sisters, x0pe02b.
6. The number of still living brothers, x0pe02c, was coded into "Missing by design" if the number of brothers, x0pe02a, was zero. Similarly, it was done for the number of living sisters, x0pe02d, and the number of sisters, x0pe02b.
7. The missing values of any questions on brothers and sisters (x0pe02a-x0pe02d) were converted into
   a) "Not in use" (-98) if the interview version, x0pever, was 1.
   b) "Unexpected missing" (-89) if the interview version was 2.
8. It was checked that no living brothers/sisters were more than the ones reported in the correspondent question x0pe02a and x0pe02b.
9. The variable on being a twin, part of a multiple birth, x0pe04, had its missing values set to
   a) "Missing by design" if no siblings were reported,
   b) "Unexpected missing" otherwise.
10. The number of offspring, x0pe11, when missing, was computed summing the number of sons, x0pe11a, and the number of daughters, x0pe11b.
11. The missing values of any questions on sons and daughters (x0pe11a-x0pe11b) were converted into
    c) "Not in use" (-98) if the interview version was 1.
    d) "Unexpected missing" (-89) if the interview version was 2.
12. The categories of the variable on civil status, x0pe06, were standardized between the two versions. For instance, given that in the second version response options 2 and 3 were collapsed, the number 2 stood for "separated/divorced".
13. The missing values of the variable on the number of persons living in the household, x0pe07, were converted into

a) "Not in use" (-98) if the interview version, x0pever, was 1.
b) "Unexpected missing" (-89) if the interview version was 2.

14. The variables on education, x0pe08 and x0pe09, had their missing values set to "Unexpected missing".

15. The variable on the ability to read or write, x0pe10, was asked to those who reported no formal education (x0pe08=1) and less than six years of school attended (x0pe09<6), only with the first version. Its missing values were set to
    a) "Not in use" if the version was the second,
    b) "Missing by design" if the version was the first and at least 6 years of school had been attended or some education degree had been attained (x0pe08>1),
    c) "Unexpected missing" if the version was the first and less than six years of school had been attended (x0pe09<6) and no formal education was reported (x0pe08=1).

16. The baseline dataset was saved.

4. Advices for the analysis

The analyst, when taking into account geographical effects, would better benefit of the use of the place of origin collected at the interview (x0pe05d and x0pe05e), rather than the birthplace as reported on the voters register (x0_birthp). Indeed, the self-reported place of origin is more important in terms of exposure and reflects more the location where the participant grew up, whereas the birthplace is more dependent on hospital location, usually occurs in larger cities, and depends also on circumstances.

Finally, the analyst should always take into account that the operator in charge of carrying out the interview might have influenced how the participant reported their answers. The analyst should therefore adjust for the operator variable, x0_opintc, when possible.

5. References

Italian education system: https://eacea.ec.europa.eu/national-policies/eurydice/content/italy_en