CHRIS Study

Interview – Birth data

Version 1.2

24th April 2024

Authors: LB, MG

1. Introduction

This module stores information related to the birth of the participants, that were collected at the interview, such as pregnancy week at delivery, birthweight, breastfeeding duration, and congenital malformations. In the invitation letter, participants are required to bring at the study center the information on their birthweight.

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 between November 5th, 2012 and November 20th, 2013. Version 3 has been in use since November 21st, 2013.

Between the different versions, the following changes have occurred:

Version 1 to Version 2:

question rephrased: x0bi05 (DE, IT), x0bi06a (DE).

Version 2 to Version 3:

question rephrased: x0bi01a (DE, IT), x0bi02a (DE).

other: x0bi03 (new response option "I do not know"), x0bi05 (changes in response option, see table 1 below).

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, except for x0bi06a, that was changed only in its German version.

3. Data cleaning

1. The main CHRIS dataset was loaded.
2. The variables with further comments were all translated, x0bin1, x0bin2, x0bin3, and x0binote, and categorized when possible.

3. The missing observations of the variable “knowing exact birth weight”, x0bi01a, were set to “Unexpected missing” (-89).
4. The missing observations of the birthweight variable, x0bi01, were set to:
    a) “Missing by design” (-99) if the exact birth weight was not known (x0bi01a=”No”),
    b) “Unexpected missing” (-89) otherwise.
5. The variable on preterm/post term birth, x0bi03, had its missing observations set to “Unexpected missing” and the “I do not know” answer, present only in the third version, set to “don’t know” missing category (-88).
6. The variable on knowing the exact pregnancy duration, x0bi02a, had its missing observations set to:
    a) “Missing by design” if the birth was normal in duration (x0bi03=”Normal”)
    b) “Unexpected missing” otherwise.

    Its third answer option “I don’t know” was turned into the missing category “Don’t know” (-88).

7. The variable of the pregnancy duration in weeks, x0bi02, had its missing observations set to
    a) “Missing by design” if x0bi02a=”Missing by design”,
    b) “Don’t know” if x0bi02a=”Don’t know”,
    c) “Unexpected missing otherwise”.
8. The distribution of pregnancy duration, x0bi02, was checked for each x0bi03 value. Consistency between preterm/post term, x0bi03, and pregnancy duration, x0bi02, were checked. There are 23 cases were pregnancy duration in weeks and preterm/post term reported do not match.
9. The variable on the type of birth, x0bi04, had its missing observations set to “Unexpected missing” and the “I do not know” answer, present only in the third version, set to “don’t know” missing category (-88).
10. The single question on breastfeeding duration, was split into two variables, one on breastfeeding yes/no and the second one breastfeeding duration, named x0bi05 and x0bi05a, respectively. The dichotomous variable, x0bi05, had options “Yes”, “No”, “Don’t know” (-88), and “Unexpected missing” (-89). The breastfeeding duration variable, x0bi05a, had options “0-2 weeks”, “3-4 weeks”, “2-3 months”, “4-6 months”, “More than 6 months”. Its missing values were either “Missing by design” (if x0bi05=”No”), or “Don’t know”, and “Unexpected missing” otherwise.
11. The variable on the presence of congenital malformations, x0bi06, had its observations set to:
    a) “Unexpected missing” if it was missing,
    b) “Don’t know” (-88) if the option “I do not know” was selected.
12. The variable on the actual type of congenital malformations, x0bi06a, with 550 participants reporting something, was translated and categorized when possible.
13. The baseline dataset was saved.

4. Advices for the analysis

The analyst should trust more the variable on the known status of birth (preterm/post term) rather than the actual pregnancy duration in weeks, more prone to recall bias and typos. Still, if the birth weight is particularly low, a preterm can be assumed if the weeks are less than 36 and post term is reported. Similarly, very high birth weights might be incompatible with a reported preterm.

Furthermore, 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.