CHRIS Study

Interview – Smoking

Version 1.1

24th April 2024

Authors: LB, MG, RM

1. Introduction

This module stores information related to smoking, that was collected at the interview, such as the frequency of smoking, the tobacco products and their daily consumption and the age at which smoking began.

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. The first version of this module is based on the RAUCHEN INKLUSIV PASSIVRAUCHEN module of the follow-up F4 questionnaire of the KORA Study (*Kooperative Gesundheitsforschung in der Region Augsburg*). Since Version 2, the module is based on the European Community Respiratory Health Research (ECRHS III) questionnaire.

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 January 21st, 2013. Version 3 has been in use since January 22nd, 2013.

Between the different versions, the following changes have occurred:

Version 1 to Version 2:

variables dropped: x0sm01b x0sm02 x0sm03 x0sm04 x0sm05 x0sm06b x0sm07 x0sm07b x0sm08 x0sm10 x0sm11b x0sm12b x0sm12c x0sm13 x0sm14 x0sm15b x0sm16b x0sm16c x0sm17 x0sm18 x0sm19b x0sm20b x0sm20c x0sm21 x0sm22 x0sm23b x0sm24b x0sm24c x0sm25 x0sm26 x0sm26a x0sm26b x0sm27 x0sm27a x0sm27b x0sm28 x0sm28a x0sm28b x0sm29 x0sm29a x0sm29b x0sm30 x0sm30a x0sm30b x0sm31 x0sm31a x0sm31b

variables added: x0sm32 x0sm33 x0sm34 x0sm34a x0sm35 x0sm36 x0sm37 x0sm38 x0sm39 x0sm39a x0sm40 x0sm41 x0sm42 x0sm43 x0sm44 x0sm45 x0sm46 x0sm47 x0sm48 x0sm49 x0sm49a x0sm49b x0sm49c x0sm49d

question order changed: x0sm00 (before: first question, then: after x0sm34a)

question filtering criteria changed: x0sm00 (before: no restriction, then: x0sm32=1)

Version 2 to Version 3:

question filtering criteria changed: error correction: before x0sm39 was asked if x0sm00=1, then if x0sm32=1

The cleaning process has added the variables x0sm51, x0sm52, x0sm53, x0sm54, x0sm55, x0sm56, x0sm57, x0sm58, x0sm59, x0sm60, and x0sm61.

3. Data cleaning

1. The main CHRIS dataset was loaded.
2. The variables on current smoking status, x0sm34 and x0sm35, were corrected based on the content of the x0smn2 note variable. When 4-10 cigarettes per week were reported, one cigarette per day was assigned; with 11-17 cigarettes per week reported, two cigarettes per day were assigned; with 18-24 per week, three cigarettes per day were assigned.
3. The same was done on the variable on past smoking status, x0sm41, using the information in the notes x0smn3.
4. If any temporary interruption of smoking was reported in x0smn2 or x0smn3 or x0smnote, this was saved in new variables representing the duration of the interruption, tmp_dur, the age at the interruption start, tmp_start, and the age at the interruption end, tmp_end. The duration of the interruption is not necessarily the difference between the end and the start, since sometimes the participant had interrupted and restarted to smoke multiple times.
5. The variable on ever smoking for at least one year, x0sm32, had its missing observations transformed into:
   a) "Not in use" (-98) if the questionnaire version, x0smver, was the first,
   b) "Unexpected missing" (-89) otherwise.
6. The variable on current cigarette smoking, x0sm00, had its missing observations transformed into:
   a) "Missing by design" (-99) if the questionnaire version, x0smver, was not the first and the ever smoking, x0sm32, was "Missing by design" or "No",
   b) "Unexpected missing" (-89) otherwise.
7. All the variables of the first version of the questionnaire, x0sm01b-x0sm08 x0sm10-x0sm25b x0sm26-x0sm31b, had their missing observations transformed into "Not in use" if the questionnaire version, x0smver, was not the first.
8. If the age of start of the current smoking habit, x0sm01b, was provided and the current smoking answer, x0sm00, was "No", the latter was transformed into "Yes".
9. Both the variable on age at cigarette start, x0sm01b, and the one on current regular or occasional smoking, x0sm02, had their missing observations transformed into
   a) "Missing by design" (-99) if the participant was not smoking at the time (x0sm00="No"),
   b) "Unexpected missing" (-89) otherwise.
10. Both the variable on daily cigarette count, x0sm03, and the one on trying to stop smoking, x0sm04, had their missing observations transformed into
   a) "Missing by design" (-99) if the participant was not regularly or occasionally smoking (x0sm02="No" or x0sm02="Missing by design"),
   b) "Unexpected missing" (-89) otherwise.
11. The variable on ever cigarette smoking, x0sm05, had its missing observations transformed into "Missing by design" if the participant was smoking cigarette at the time (x0sm00="Yes").

12. Both the variable on age at past smoking start, x0sm06b, and the one on timing at smoking stop, x0sm07, had their missing observations transformed into
    a) "Missing by design" if the participant never smoked cigarettes (x0sm05="No" or x0sm05="Missing by design"),
    b) "Unexpected missing" otherwise.
13. The categorical variable on the past smoking stop, x0sm07, was further corrected into "More than one year ago" when it was missing and the age or the year at smoking stop, x0sm07a and x0sm07b, was given.
14. The variable on the age at smoking stop, x0sm07b, had its missing observations transformed into "Missing by design" (-99) if the categorical variable on the smoking stop, x0sm07, was "Missing by design" or less than a year before.
15. The variable on past average cigarette daily count, x0sm08, had its missing observations transformed into "Missing by design" if the participant never smoked cigarettes (x0sm05="Missing by design" or "No").
16. The remainder missing observations of x0sm07b and x0sm08 were transformed into "Unexpected missing".
17. The variables on pipe smoking ever, x0sm10, cigar smoking ever, x0sm14, tobacco chewing ever, x0sm18, and tobacco sniffing ever, x0sm22, were transformed as follows:
    a) "Unexpected missing" (-89) if they were missing,
    b) "Don't know" (-88) if the third answer option "Don't know" was chosen.
18. The variable on age at pipe smoking start, x0sm11b, on age at pipe smoking stop, x0sm12b, and on current pipe smoking, x0sm12c, had their missing observations transformed into "Missing by design" if the participant had never smoked pipe (x0sm10="No" or "Missing by design")
19. The variable on age at pipe smoking start, x0sm11b, had the remaining missing observations set to "Unexpected missing".
20. The variable on age at pipe smoking stop, x0sm12b, was changed into the participant's age, x0_age, if the participant was still smoking the pipe.
21. The variable on current pipe smoking, x0sm12c, had its missing observations set to "Missing by design" if an age at pipe smoking stop, x0sm12b, was provided.
22. The remaining missing observations of x0sm12b and x0sm12c were set to "Unexpected missing".
23. The daily and monthly pipe smoking count, x0sm13a and x0sm13b, were merged into a single variable, x0sm13, which was equal to the monthly count, when provided, and to the daily count multiplied by 30, when the daily count was provided. This new variable had its missing observations transformed into:
    a) "Not in use" if the version was not the first,
    b) "Missing by design" if the participant never smoked the pipe (x0sm10="No" or "Missing by design" or "Don't know"),
    c) "Unexpected missing" otherwise.
24. The original daily and monthly pipe count variables, x0sm13a and x0sm13b, were dropped.
25. The steps 17-24 of the cleaning process were repeated for the cigar smoking variables, x0sm14 to x0sm17b. The new variable on monthly cigar amount was called x0sm17.
26. The steps 17-24 of the cleaning process were repeated for the tobacco chewing variables, x0sm18 to x0sm21b. The new variable on monthly tobacco chewing amount was called x0sm21.

27. The steps 17-24 of the cleaning process were repeated for the tobacco sniffing variables, x0sm22 to x0sm25b. The new variable on monthly tobacco sniffing amount was called x0sm25.

28. All the variables related to a year of start/stop of a smoking habit, x0sm01a, x0sm06a, x0sm07a, x0sm11a, x0sm12a, x0sm15a, x0sm16a, x0sm19a, x0sm20a, x0sm23a, and x0sm24a, were dropped in favor of the correspondent age variables.

29. All the variables introduced with version 2, see table 2 below, i.e. x0sm33-x0sm49d, had their missing observations set to "Not in use" if the questionnaire version, x0smver, was the first.

30. The variables on the age at smoking start (no specification of which tobacco product), x0sm33, and at daily smoking start, x0sm34, had its missing observations set to:
    a) "Missing by design" if the participant never smoked for a full year (x0sm32="No" or "Missing by design"),
    b) "Missing by design" if the participant never smoked daily (only for x0sm34: x0sm34a="No"),
    c) "Unexpected missing" otherwise,
    d) "Don't know" if the observation had value 99.

31. The variable on daily smoking ever, x0sm34a, had its missing observations set to:
    a) "Missing by design" if the participant never smoked for a full year (x0sm32="No" or "Missing by design"),
    b) "Don't know" if the age provided, x0sm34, was 99,
    c) "Never smoked daily" if the age provided was 88,
    d) "No" if an age was provided that was neither 88 nor 99.

32. The variable on ever smoking daily was corrected from "Never smoked daily" to "No" if an age of daily smoking start was provided. The remaining missing observations of x0sm34 and x0sm34a were set to "Unexpected missing".

33. The variables on the current daily count of cigarettes, cigarillos, cigars, and grams of pipe tobacco, x0sm35-x0sm38, had their missing observations transformed into:
    a) "Missing by design" if the participant never smoked (x0sm00="No" or "Missing by design"),
    b) "Unexpected missing" otherwise.

34. If no current smoking quantity was reported but at least 1 cigarette daily or other tobacco products in positive daily amounts were reported, x0sm34a was changed from "Never smoked daily" to "No".

35. The variable on smoke reduction/stop, x0sm39, had its missing observations set to:
    a) "Missing by design" if the participant never smoked for at least a full year (x0sm32="No" or "Missing by design"),
    b) "Unexpected missing" otherwise.

36. The variables on smoking reduction/stop due to respiratory problems, on age at smoking reduction/stop, x0sm39a and x0sm40, as well as all the past tobacco daily counts and smoke inhalation, x0sm41-x0sm45, had their missing observations transformed into:
    a) "Missing by design" if the participant never reduced or stopped smoking (x0sm32="No" or "Missing by design")
    b) "Unexpected missing" otherwise.

37. The variable on age at smoking reduction/stop, x0sm40, was further corrected to "Unexpected missing" when the age reported was zero, and to "Don't know" when the age reported was 99.

38. Inconsistencies between the various age at start, daily start, reduction/stop, and age at examination were checked. When the ages at reduction/stop were lower than the age at start and below 11 years of age, they were set to "Out of range" (-86).

39. The questions on passive smoking in the household, x0sm26, x0sm27, x0sm28, had their missing observations transformed into "Missing by design" if the previous question had already a "No" or "Missing by design", and "Unexpected missing" otherwise.

40. The question on passive smoking at the workplace, x0sm29, had its missing observations set to "Unexpected missing".

41. The following questions on further periods of passive smoking at the workplace, x0sm30 and x0sm31, had their missing observations transformed into "Missing by design" if the previous question had already a "No" or "Missing by design", and "Unexpected missing" otherwise.

42. For each passive smoking variable asking for a precise year of start/end of exposure to passive smoking, they had their missing observations transformed to "Missing by design" if the corresponding question had a "No" or "Missing by design", and to "Unexpected missing" otherwise. The values 9999 were transformed to "Don't know".

43. The variable on regular passive smoking exposure in the previous 12 months, x0sm46, had its missing observation set to "Unexpected missing".

44. The variables on home and workplace passive smoking exposure and general daily exposure duration, x0sm47, x0sm48, and x0sm49, had their missing observations set to:
    a) "Missing by design" if the passive exposure in the previous 12 months was not regular (x0sm46="No" or "Missing by design"),
    b) "Unexpected missing" otherwise.

45. The variables on daily passive smoking in different places, x0sm49a-x0sm49d, had their missing observations set to:
    a) "Missing by design" if the hours of daily passive smoking exposure are zero (x0sm49=0 or "Missing by design"),
    b) "Unexpected missing" otherwise.

46. If the general daily passive smoking exposure variable, x0sm49, was missing but all the variables on daily passive smoking in different places, x0sm49a-x0sm49d, were provided, the former was substituted by the sum of x0sm49a, x0sm49b, x0sm49c, and x0sm49d.

47. A comprehensive smoke habit variable, x0sm51, was created make use of current and past smoking habits:
    a) x0sm51="Never smoker" if the participant never smoked any tobacco product (x0sm00=="No" and x0sm05=="No" and x0sm10=="No" and x0sm14=="No") or the participant never smoked for at least a whole year (x0sm32="No"),
    b) x0sm51="Past smoker" if the participant does not currently smoke cigarettes but did in the past (x0sm00="No" and x0sm05="Yes, regularly" or "Yes, occasionally"), or if the participant does not currently smoke pipe but previously did (x0sm10="Yes" and x0sm12c="Missing by design"), or if the participant does not currently smoke cigars but previously (x0sm14="Yes" and x0sm16c="Missing by design"), or the participant never smoked for a full year and does not current smoke (x0sm00="No" and x0sm32="Yes"),
    c) x0sm51="Current smoker who reduced", if x0sm00="Yes" and x0sm39="Yes",
    d) x0sm51="Current smoker who did not reduce", if x0sm00="Yes" and x0sm39="No",

e) x0sm51="Current smoker with unknown reduction status", if x0sm00="Yes" and x0sm05="Missing by design", or x0sm10="Yes" and x0sm12c="Until now", or x0sm14="Yes" and x0sm16c="Until now", or x0sm00="Yes" and x0sm39="Unexpected missing",
f) x0sm51="Ever smoker with unknown current status" if x0sm00="Unexpected missing" and x0sm32="Yes",
g) x0sm51="Unexpected missing" if x0sm32="Unexpected missing".

Note that having stopped smoking for less than the last 6 months makes the participant still a current smoker.

48. A comprehensive daily smoke amount was computed using the number of daily tobacco products and these conversion rates: Each cigarette corresponded to 1 tobacco gram, each cigarillos to 3 g, each cigar to 5 g and each pipe to 2.85 g (0.1 ounce). For the first version of the interview, no amount of tobacco contained in cigarillos could be computed. The monthly amounts were transformed into daily amounts by multiplying them by 12/365.25, the weekly amounts by dividing by 7. The total daily amount was computed summing the tobacco amount of cigarette, cigarillos, cigars, and pipes, and saved as x0sm57. The variable was transformed into "Not in use" (-98) whenever the interview version was the first, and to "Missing by design" when it was missing for past or never smokers, according to x0sm51 values. The remaining missing observation for x0sm57 were set to "Unexpected missing".

49. The same procedure was done to compute the past tobacco daily amounts, saving the results into the variable x0sm58, also available only for versions 2 and 3.

50. The lifetime cigarette smoking in pack-years, x0sm52, could be computed for participants answering the first version of the interview, considering 20 cigarettes per pack:
    a) If the participant was still smoking, then the duration was the difference between age at examination and age at current smoking start, x0sm01b; if present in the notes, the interruption period length x0sm06d was subtracted to the duration,
    b) If the participant had stopped smoking cigarettes, the duration was the difference between age at smoking stop, x0sm07b, and age at smoking start, x0sm06b; if present in the notes, the interruption period length x0sm06d was subtracted to the duration,
    c) If the participant had stopped smoking cigarettes since less than one year, 6 months' time were subtracted to the difference between age at examination and age at smoking start; if present in the notes, the interruption period length x0sm06d was subtracted to the duration,
    d) The value 0 was assigned to those still with x0sm52 missing, who never smoked cigarettes,
    e) The value "Not in use" (-98) was assigned to those answering version 2 and 3, the value "Missing by design" to those who never smoked cigarettes, and "Unexpected missing" to all the remaining missing observations.

51. The same procedure described in step 50 was used to compute the lifetime cigar smoking in pack-years, x0sm53, considering 20 cigarettes per pack.

52. The same procedure described in step 50 was used to compute the lifetime pipe smoking in pack-years, x0sm54, considering 20 cigarettes per pack.

53. Then a lifetime total smoking variable, summing up lifetime cigarette, cigar, and pipe smoking variables, x0sm52-x0sm54, whenever these were reported. This was called x0sm55.
54. For the participants answering the second or third version, the lifetime smoking based on the current smoking, x0sm59, was computed based on current smoking quantities, x0sm57, and the smoking duration. The smoking duration was either the difference between age at examination and age at smoking reduction, x0sm40, or the difference between age at examination and age at smoking start, x0sm33.
55. For the participants answering the second or third version, the lifetime smoking based on past smoking habits, x0sm60, was computed based on past smoking quantities, using as duration the difference between the year at smoking reduction, x0sm40, and the year at smoking start, x0sm33.
56. The sum of current and past lifetime smoking, x0sm59 and x0sm60, was assigned to the lifetime total smoking variable, x0sm55, for participants who answered the second or third version. For those with either the current or past lifetime smoking missing, only the other quantity was assigned to x0sm55.
57. The age at any type of smoking onset, x0sm56, was created as the minimum between the various age starts provided for the first version, or as x0sm33 for the following versions.
58. Inconsistencies between smoking habits and smoking status, x0sm51, were checked and solved
59. For those participants who started smoking in the year of the examination, packyear values in x0sm52, x0sm53, and x0sm54 were transformed from zero to the reported values multiplied by 0.5 year.
60. If the participants reported to currently smoke but had null quantities in all the smoking lifetime in pack-years, an "Unexpected missing" was assigned to x0sm51-x0sm54. The total lifetime smoking variable, x0sm55, was then corrected accordingly.
61. A new variable on daily tobacco smoking in grams/day of any type of tobacco product, only for current smokers, x0sm61, was created using the variables x0sm03, x0sm13, and x0sm17 for version 1 and copying x0sm57 for versions 2 and 3.
62. The baseline dataset was saved.

4. Advices for the analysis

The content of the nurse's notes, referring to smoking breaks or to occasional smoking habits, have already been included in the analysis, so the variable was not made available.

The retrospective nature of this assessment makes the measurement of alcohol consumption habits prone to recall bias. Furthermore, social desirability bias might have pushed participants to understate their smoking exposure, especially given the questions are asked by a study nurse.

If the analyst is interested in the age at onset of smoking, they'd better use the variable x0sm56, as it is based on the minimum plausible age reported. Indeed, in the first version, participants could report a different onset age for each tobacco product, and in the following versions they could also report different onset age for smoking and regular smoking.

In the variable on smoke habit, x0sm51, the options "Current smoker who reduced" and "Current smoker who did not reduce" are only possible for versions 2 and 3 of the survey. Instead, "Current smoker with unknown reduction status" is possible only with version 1.

Some derived variables are available only for the first version, such as x0sm52-x0sm54, or for the second and third one, such as x0sm57-x0sm60, whereas the ones that are available for all versions are x0sm51, x0sm55, x0sm60, and x0sm61.

Exposure to passive smoking in terms of hours is difficult to estimate and the daily duration for each location (x0sm49a-x0sm49d) might not add up to the general passive smoking exposure variable x0sm49.

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

Janson C, Anto J, Burney P, Chinn S, de Marco R, Heinrich J, Jarvis D, Kuenzli N, Leynaert B, Luczynska C, Neukirch F, Svanes C, Sunyer J, Wjst M. European Community Respiratory Health Survey II. The European Community Respiratory Health Survey: what are the main results so far? European Community Respiratory Health Survey II. Eur Respir J. 2001 Sep;18(3):598-611. DOI: 10.1183/09031936.01.00205801

Löwel H, Döring A, Schneider A, Heier M, Thorand B, Meisinger C. The MONICA Augsburg surveys - basis for prospective cohort studies. Gesundheitswesen. 2005;67(Sonderheft 1):S13–S18. DOI: 10.1055/s-2005-858234

Holle R, Happich M, Löwel H, Wichmann H-E. KORA-A Research Platform for Population Based Health Research. Gesundheitswesen. 2005;67(Sonderheft 1):S19–S25. DOI: 10.1055/s-2005-858235

Murgia F, Melotti R, Foco L, Gögele M, Meraviglia V, Motta B, Steger A, Toifl M, Sinnecker D, Müller A, Merati G, Schmidt G, Rossini A , Pramstaller PP, Pattaro C. Effect of smoking status, history and intensity on heart rate variability in the general population: The CHRIS study, PLoS One 2019 Apr 9;14(4):e0215053. DOI: 10.1371/journal.pone.0215053

ECHRS III questionnaire: https://www.ecrhs.org/ecrhs-iii