

# **CHRIS Study**

## **Interview – Occupation**

Version 1.1

24<sup>th</sup> April 2024

Authors: LB, MG

## 1. Introduction

This module stores information related to the current and past occupation was retrieved in the CHRIS study with 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.

The module on occupation is based on the questionnaire of the KORA study (*Kooperative Gesundheitsforschung in der Region Augsburg*) follow-up F4 questionnaire, module SOZIODEMOGRAPHIE UND BERUF.

## 2. History of the versions

Between the different versions, the following changes have occurred:

### Version 1 to Version 2:

**variables dropped:** x0oc12 (unemployment benefits in the last 2 years)

### Version 2 to Version 3:

**variables added:** x0oc02e (first occupation, until now?); x0oc03e (second occupation, until now?); x0oc04e (third occupation, until now?); x0oc05e (fourth occupation, until now?); x0oc06e (fifth occupation, until now?); x0oc07e (sixth occupation, until now?)

## 3. Data cleaning

### Whole module cleaning:

1. The variable x0oc01 was recoded from the original formulation "Have you been employed or self-employed before?" (1 yes/ 2 no) into "Employment ever", with answer levels (1 Yes, in the past, 2 Yes, currently, 3 No), based on the answers of x0oc00.
2. If the only occupation provided in x0oc02a and x0oc2b was not a profession, e.g. "housewife", "student", "civil servant", and the variables x0oc00 and x0oc01 indicated the participant was ever employed, both the variables x0oc00 and x0oc01 were assigned the answer "No" (levels 4 and 3, respectively). This change was performed for 17 participants.
3. The question "Until when have you been regularly employed?", x0oc01a, has been recoded as:
  - a) -99 (missing by design) if x0oc01="Yes, in the past" OR x0oc01= "Yes, currently";
  - b) -89 (unexpected missing) if x0oc01=="No" or x0oc01=-89,

- c) -88 (don't know) if x0oc01a==9999. Interviewers were instructed to insert 9999 when the participant did not know the answer.
- 4. The missing values in x0oc10 were recoded into
  - a) "Pensioner" if x0oc01= "Yes, in the past" OR "No" and x0\_age>=75
  - b) -99 (missing by design) if x0oc01=="Yes, currently"
  - c) -89 otherwise
- 5. If in the free text answers, the current position was "high school student", the variable x0oc10 (status if not employed) was switched from "in education and retraining" to "student".
- 6. For those participants described in step 2 that wrongfully declared to have a profession, even if theirs is not one, were assigned in x0oc10 the values of "student", "housewife/househusband", and "in the military service or alternative service", according to their x0oc02b answer.
- 7. Two participants, one declaring invalidity, and one employed in the past and older than 65, were both assigned in x0oc10 the value "pensioner/retiree". Another participant, a full time caregiver, was assigned the value "In maternity protection, parental leave or other leave". From the notes of the interviewer, an additional participant has taken time off work, this was assigned "unemployed" in x0oc10.
- 8. The variable x0oc03 (Second occupation: another occupation?) has its missing values set to -99 (missing by design) if the participant has never been employed (x0oc01="No"). Then the variable x0oc04 (Third occupation: another occupation?) has the missing values set to -99 if x0oc03="No" or x0oc03="missing by design". Similarly, x0oc05, x0oc06, and x0oc07 have their missing values replaced by -99 in case the previous occupation is already missing.
- 9. The variable x0oc02a (First occupation: list) has its missing values changed to:
  - a) "other" if x0oc02b has some text,
  - b) "missing by design" if x0oc01="No"
  - c) "unexpected missing" otherwise
- 10. The variables x0oc03a has their missing values changed to:
  - a) "other" if x0oc03b has some text,
  - b) "missing by design" if x0oc03="No" or x0oc03="missing by design".
  - c) "unexpected missing" otherwise

The same procedure occurs for x0oc04a, x0oc05a, x0oc06a, and x0oc07a.

- 11. The variable x0oc02a is transformed into one of its categories if the text in x0oc02b points to an already suggested professional category.
- 12. The years of start and end of the first occupation, x0oc02c and x0oc02d, have their missing values changed:
  - a) "missing by design" if x0oc02a="missing by design"
  - b) X0oc02d=x0\_examy (year of examination) if it is missing and they stated they worked until now (x0oc02e="Yes").
  - c) "unexpected missing" otherwise.

The same procedure applies to x0oc03c, x0oc03d, x0oc04c, x0oc04d, x0oc05c, x0oc05d, x0oc06c, x0oc06d, x0oc07c, x0oc07d.

13. The variable on the physical activity intensity of the regular occupation, x0oc08, has its missing values set to:
  - a) “missing by design” if the participant has never worked (x0oc01=“No”) OR if the only reported occupation is not a working position, i.e. “housewife”, “student” or “civil servant”,
  - b) “unexpected missing” if the participant currently works or worked in the past (x0oc00=“Yes” OR x0oc01=“Yes, in the past”)
14. The variables x0oc11a and x0oc11b, respectively the month and year of the unemployment start, are completed if there are additional notes from the interviewer. Furthermore, its missing values are set to:
  - a) “missing by design” if the participant is not unemployed (x0oc10!=“unemployed”)
  - b) “unexpected missing” otherwise
  - c) If the month variable x0oc11a is available and x0oc11b is missing, the question was asked as “Since how many months are you unemployed?”, and reported in the interviewer’s notes. Therefore if x0oc11a was smaller than examination month (x0\_examm), x0oc11b was set to the examination year x0\_examy and x0oc11a to x0\_examm-x0oc11a. If instead x0oc11a was higher than examination month (x0\_examm), x0oc11b was set to the year previous to the examination (x0\_examy -1) and x0oc11a to x0\_12+ examm-x0oc11a.
15. The variable x0oc12 on unemployment benefits was in use only for version1. Therefore. its missing values were coded as follows:
  - a) “not in use” (-98) if x0ocver!=1
  - b) “missing by design” if x0age>75 OR the participant is not in the work force (x0oc00=“No”) OR if x0oc10 is missing (i.e. the participant has not explicitly stated they are unemployed or on some leave).
  - c) “unexpected missing” otherwise
16. The variable x0oc13, about the receipt of an invalidity pension, has its missing values set as:
  - a) “missing by design” if x0age>75 or the participant is not in the work force (x0oc00=“No”) or if x0oc10 is missing (i.e. the participant has not explicitly stated they are unemployed or on some leave).
  - b) “unexpected missing” otherwise.
17. The variables regarding the duration of invalidity pension, x0oc13a (start year), x0oc13b (end year), and x0oc13c (until now) have been corrected as follows:
  - a) The variable x0oc13b has been replaced by x0\_examy if missing and x0oc13c=“Yes”
  - b) Both start and end year have been set to “missing by design” if x0oc13=“No” or if x0oc13 was “missing by design”.
  - c) Missing values of x0oc13a and x0oc13b have been replaced by “unexpected missing” otherwise.

#### **Classification process:**

1. If more than one job was mentioned in the same free text variable, e.g. x0oc02b, the professions were split into two distinguished professions, with the same start and end year.

2. Only the variables aid, x0oc00-x0oc07e were kept. The dataset was reshaped into long form, so that each row corresponded to a single work experience of a participant. The profession position was stored in a variable called job\_number.
3. The free text variable x0oc02b was standardized grouping the answers describing the same profession in terms of level and sector. The standardized variable is called x0oc02b\_std
4. The frequency of each mentioned profession was assessed. Out of 21,110 professional experiences, around 10,795 consisted of the categories listed in x0oc02a. The remaining 10,315 professions were grouped into additional 1949 different professions.
5. Three researchers classified all professions with a frequency of at least four (N=350). They assigned a professional level based on type of activity and education needed (nine major groups of the Classificazione delle professioni CP2011 scale, from the Italian National Institute of Statistics, ISTAT); as well as a sector based on the 21 main categories of the Statistical Classification of European Classification of Economic Activities (NACE) Rev 2. As of 2011, ISTAT adopted the CP2011 classification of occupations, which is the result of updating the previous version (CP2001) and adapting it to the changes introduced by the International Standard Classification of Occupations – ISCO-08.

The CP2011 classification describes all the existing occupations in the labour market with a limited number of occupational groupings, and it is meant be used to communicate and exchange statistical and administrative data on occupations, not to regulate occupations.

The object of the classification, the profession, is defined as a set of work activities concretely carried out by an individual, involving knowledge, skills, identity, and status.

The logic used to aggregate different professions is based on the concept of competence, seen both as the level and the field of competence required to exercise that profession.

The level of competence is defined according to the complexity, the extent of the tasks performed, the level of responsibility and decision-making autonomy that characterises the profession; the field of competence captures, instead, the differences in sectoral domains, in the disciplinary fields of applied knowledge, in the equipment used, in the materials processed, in the type of goods produced or service provided.

The competence criterion outlines a classification system articulated on 5 hierarchical levels of aggregation: the first level consisting of 9 major professional groups; the second level, with 37 professional groups; up to the fifth and last level, with 800 occupational units, within which the professions existing in the labour market can be traced.

When the researcher could not classify precisely the major professional group, the instruction was to assign the specifically added tenth category, namely “not classifiable”.

NACE is the European reference framework for data regarding economic activities. The scale NACE was developed in 1970, further revised until the most recent version NACE Rev 2 was adopted in 2006. Its creation stems from a discussion coordinated by Eurostat which involved also the various EU national statistical institutes, European trade and business associations, the European Central Bank, and the United Nations Statistical Division. An economic activity is characterised by an input of resources, a production process and an output of products (goods or services). NACE does not distinguish activities based on the ownership of a production unit or its legal organisation or mode of operation, neither does it draw distinctions based on legal or illegal/formal or informal production. NACE also consists of a hierarchical structure, whose first level is identified by 21 headings identified by alphabetical letters. At this level, also called

sections, the general features of the goods and services produced, as well as the potential use of the statistics, become an important factor.

When the researcher could not uniquely classify the major sector of activity, the instruction was to assign the specifically added 22<sup>nd</sup> category, namely “not classifiable”. Furthermore, when an activity was carried out in a specific environment/institution, the sector to assign was the one of the institution: for instance if someone reported to perform cleaning activities in an hospital, the sector assigned was “Q Human health and social work activities”.

6. Two researchers also classified the remaining professions with frequency below four (N=1609), both on the CP2011 and the NACE Rev 2 scale.
7. The two researchers discussed on all the professions where their classifications differed and reached an agreement on a combined classification.
8. The two researchers created macro-categories of profession levels, for those “not classifiable” cases on the CP2011 scale. An example is the text “Arbeiter” (worker), which is not uniquely classifiable into one of the nine CP 2011 levels, but rather could belong to the categories ranging from 6 to 8. Similarly, “Angestellt”, which means both employed and clerk, was assigned to the macro-category ranging from 3 to 4.
9. Other variables that were created: a text note for the “not classifiable” professions x0oc02h, , a binary variable to identify seasonal professions x0oc02l.
10. An ID to each x0oc02b\_std value was assigned, to create a “dictionary file”, to then be merged with the occupational module dataset.
11. The resulting dataset was returned into wide format, exploiting the participant study number and the job\_number variables’ combination, with one row per participant.
12. The entire occupational classification dataset was merged back to the interview dataset, using then the x0oc02b\_std text, with a single merge.

#### **4. Advices for the analysis**

To treat the level of profession, please refer to the variables x0oc02f-x0oc07f, where the Italian version of ISCO-08 is assigned to each profession, since the categories in x0oc02a-x0oc07a do not correspond to a specific scale of professional activity, and around half of the professions belonged to the “Other” category.

In any appraisal of this module, please always consider the operator variable, as this module is part of an interview, and it is shown that operator can have an effect.

#### **5. References**

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](https://doi.org/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](https://doi.org/10.1055/s-2005-858235)

ISCO-08: <https://www.ilo.org/public/english/bureau/stat/isco/docs/publication08.pdf>

CP 2011 navigation system (in Italian): <http://professioni.istat.it/sistemainformativoprofessionioni/cp2011/>

CP 2011 description (in Italian): <https://www.istat.it/en/archivio/18421>

NACE introduction:

<https://ec.europa.eu/eurostat/documents/1965800/1978839/NACEREV.2INTRODUCTORYGUIDELINESE N.pdf/f48c8a50-feb1-4227-8fe0-935b58a0a332>

NACE Rev 2 full classification: <https://ec.europa.eu/eurostat/documents/3859598/5902521/KS-RA-07-015-EN.PDF>