



Description of the «Public use file» (PUF) for the Swiss National Cohort (SNC) dataset

Neuchâtel, 2022

Published by: Federal Statistical Office (FSO)

Information: info.dem@bfs.admin.ch

Contents: Claudia Berlin ISPM, University of Bern

Topic: 01 Population

Original text: German

Translation: FSO language services

Layout concept: Section DIAM

Downloads: www.statistics.ch

Copyright: FSO, Neuchâtel **2022**
Reproduction with mention of source authorised
(except for commercial purposes)

Description of the "Public use file" (PUF) for the SNC dataset

The Swiss National Cohort (SNC) is a cohort based on Switzerland's population. The SNC is based on census data from the years 1990 and 2000, as well as on register-based census information that was linked until 2018 with the cause of death statistics and emigration data (amongst others). Due to a lack of personal indicators in the respective data sets prior to 2010, probabilistic linkage was used to connect the data sets. Since 2010, the data sets used contain the AHV/AVS insurance number (AHVN13) (in pseudonymic form), which is used to connect the different data sets.

The SNC's main data set currently comprises more than 12 million persons and contains a large amount of information on each person from various FSO data sources, such as the population census, STATPOP and BEVNAT. The main areas are personal details (e.g. sex, date of birth, marital status, nationality, education, religion, number of persons in household, and commune of residence), and in the case of deceased persons, details such as the date and cause of death and comorbidities. Many variables are repeated, as they are available for each survey period but can change over time. Other variables such as education and religion, for example, are not always recorded and are therefore only available for certain survey periods. Other information indicates how long a person has been included in the cohort or when they left it.

The SNC's main data set contains only a selection of the variables available. Additional variables can be added from the population census (1990 and 2000), the register-based census information (available for every year since 2010), the Federal Register of Buildings and Dwellings, the statistics on live births, stillbirths, marriages, divorces and the Cause of death statistics. Information from the structural survey as of 2011 can be used to complete details for a portion of people in the SNC's main data set.

The aim of the public use file is to give a first impression of the SNC data set. It contains a selection of just 9 variables. The variable categories have been partially changed in order to exclude any conclusions about individual persons. For example, information about the precise cause of death is summarised in only 3 categories and the age is used instead of the date of birth. We have used the population census 2000 as reference data for the public use file. Thus the statistical population contains only the 7 280 246 people who were recorded in the 2000 census and are contained in the SNC main data set.

Description of variables

- id: Anonymous personal number
- age_start: Age at entry into cohort
- age_stop: Age at exit from cohort (through death, emigration or end of observation period)
- sex: Gender
- v0_civil: Marital status in 2000
- v0_nat_bin: Nationality (category) in 2000
- gross_reg: Place of residence by major region in 2000
- v0_urban: Urbanisation level of place of residence in 2000
- cause_death: Cause of death