CHRIS Study

Personal data

Version 1.1

24th April 2024

Authors: LB, MG

1. Introduction

The personal information of each participant, such as date of birth, sex, place of birth, and place of residence, was collected from the voter registers, which were also the source for the study recruitment, that are stored in the archives of each of the 13 municipalities involved.

2. History version changes

The cleaning process resulted in the following variables added:

variables added: x0_age, x0_ager, x0_birthm, x0_birthy.

3. Data cleaning

1. The main dataset was loaded.
2. The rows were sorted by order of appointment date.
3. Some inconsistency was present between the variables x0_sex, stemming from the interview, and the gender variable, stemming from the voter registers. However, the gender variable was considered more reliable, therefore x0_sex was dropped, and the gender variable was renamed as x0_sex.
4. The variable birth_date, in string format, was transformed into date format yyyy-mm-dd and saved into the variable x0_birthd. Due to reidentification risk, this variable cannot be shared.
5. From the variable x0_birthd, the month and year of birth variables were created: x0_birthm and x0_birthy.
6. The variable appointment, in string format, was transformed into date format yyyy-mm-dd and saved into the variable x0_examd.
7. From the variable x0_examd, the month and year of the examination variables were created: x0_examm and x0_examy.
8. The variable x0pe01 on the birthdate reported at the interview (see document on the Interview-personal data) was converted into date format yyyy-mm-dd and saved into x0pe01a.
9. The variable x0_birthd, stemming from the voter registers, and the x0pe01a, reported at the interview and written down by the study nurse, were compared when they did not coincide. Either the reported date was missing, or some typo occurred. Each was manually checked with their tax code, from which the birthdate can be derived. The date originating from the voter registers was always the correct one. The reported variable, x0pe01a, was therefore dropped.
10. The age of the participant at the examination date was computed as the difference between x0_birthd and x0_examd, divided by 365.25, and stored in the variable age, which is not available due to reidentification risk. Then it was renamed as x0_age.
11. The rounded age of the participant at the examination date was computed as the difference between x0_birthd and x0_examd, divided by 365.25, then rounded to the closest integer, and stored in the variable x0_ager.
12. The variable birthplace was renamed as x0_birthp.
13. The variable municipality, referring to the place of residence, was renamed as x0_residp.
14. The variable place of birth, x0_birthp, as stored in the voter register, was standardized and if only the name of the city was cited, its country was added. Due to reidentification risk, also this variable is not available.
15. The variable x0_birthp was further categorized into six groups:

a) Vinschgau district
b) South Tyrol (except Vinschgau)
c) Italy (except South Tyrol)
d) German-speaking countries
e) Other European countries
f) Non-European countries

16. For persons who moved from Vinschgau to other places, the variable x0_residp was set back to their original municipality (as in the voter register).

4. Advices for the analysis

The analyst, when taking into account geographical effects, would better benefit of the use of the origin place 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.

5. References

Pattaro C, Gögele M, Mascalzoni D, Melotti R, Schwienbacher C, De Grandi A, et al. The Cooperative Health Research in South Tyrol (CHRIS) study: rationale, objectives, and preliminary results. J Transl Med. 2015;13:348. DOI: 10.1186/s12967-015-0704-9