CHRIS Covid-19 Study

General information

Version 1.0

19 May 2022

Authors: MG

1. Introduction

This module stores administrative information related to the participant, the household identifier, a summary of the available CHRIS COVID-19 data (longitudinal questionnaires, biochemical traits, vaccination status), and data related to the random sample study.

The CHRIS COVID-19 study was designed to estimate the distribution of SARS-CoV-2 infection cases in Val Venosta/Vinschgau since 1 February 2020, as well as the proportion of asymptomatic individuals among positive cases, to characterize transmission within households, to assess the relationship between antibody response and disease severity, to observe the evolution of antibody response over time, to identify environmental, molecular and genetic risk factors, and to identify long-term sequelae.

The CHRIS COVID-19 study was organized in three stages:

Stage 1: A stratified random sample of 1812 CHRIS study participants was selected to represent the adult population of the Val Venosta/Vinschgau district. Out of this sample, 845 CHRIS participants replied to an online or paper questionnaire, underwent a molecular test based on a nasopharyngeal swab and a serum antibody test.

Stage 2: All 13,393 CHRIS study participants and their consenting cohabitants were invited to fill in an online questionnaire on their past and current health status, and on SARS-CoV-2 (potential) exposure and testing. Each CHRIS participant received ten access tokens to let them and their cohabitants register online. A shorter questionnaire was then sent repeatedly to all participants every 4 weeks for an update on their symptoms, COVID-19 exposure, and testing. All individuals at risk of positivity to SARS-CoV-2 infection and their cohabitants have been invited for a nasopharyngeal swab molecular test and a serum antibody test at the CHRIS study center.

Stage 3: To trace and monitor antibody response over time, all individuals testing positive to either the nasopharyngeal or the serum test in Stages 1 or 2 have been invited to repeat the serum antibody test every three months for a year, since their first measurement.

2. History version changes

Data cleaning process

variables added: all variables were based on administrative information or were created during the data cleaning process

3. Data cleaning

1. The study registration form dataset was loaded in Stata.
2. The data availability indicators cc_baseinfo, cc_biochem, cc_longinfo, and cc_vaccine were created by merging the study registration form dataset with the datasets containing the cleaned baseline information, biochemical traits, longitudinal information, and vaccination status variables, respectively.
3. The residential municipality variable ccbq02, the number of household members variable ccbq03, and the total number of rooms in the house variable ccbq04 were imported from the baseline information dataset.

4. The CHRIS baseline Participant ID of the token used for the study registration was added to the data set. A unique address for the CHRIS baseline participants was created and used to impute the address of co-inhabitants who used a token associated to the CHRIS baseline participant.
5. For CHRIS COVID-19 participants without address information, the information was retrieved from the CHRIS person manager database. Afterwards, the imputation process described in point 4 was repeated. It was possible that only the co-habitants of a CHRIS participant took part in the CHRIS COVID-19 study but not the CHRIS baseline participant living with them. Additionally, also household members younger than 18 years old could participate. It was possible for any CHRIS COVID-19 participant to fill in a questionnaire for another household member (variable cclqresp).
6. A unique household identifier cc_hhid was associated to all individuals with the same address.
7. A new number of inhabitants variable ccbq03a and total number of rooms variable ccbq04a was created. The median of the valid responses was used to impute missing responses. In an iterative process, in case the number of household members according to the household identifier was larger than the number reported in the baseline questionnaire, the personal data were reviewed, and the participants were manually divided in two or more households, as appropriate.
8. The variables ccbq03a and ccbq04a were re-estimated. A residential municipality variable ccbq02a was created and missing responses were imputed from the assigned address information.
9. The newly created variables ccbq02a, ccbq03a, and ccbq03a belonging to the baseline information dataset were saved in the baseline information dataset.
10. The information about the sex of the participant cc_sex was derived from the fiscal code.
11. The information about the date of birth of the participant was derived from the fiscal code and the command `personage' was used to calculate the participant's age on February 1st, 2020, as the CHRIS COVID-19 study reference date. The command also truncates the age to its integer part.
12. The information about the inclusion of CHRIS participants in the random sample selection cc_rssele was added to the dataset as well as the availability of biochemical traits collected during the random sample examination in July and August 2020, cc_rsexam.
13. For individuals who participated to the random sample examination, the expansion and relative sampling weight variables, named cc_esw and cc_rsw respectively, were added to the dataset.
14. The general information dataset was saved.

Stata v16.1 was used for the data cleaning process. The cr_09_IC_household do-file is not available because it includes personal data of the participants.

4. Data structure

The variables listed in table 1 constitute all the variables related in general to the CHRIS COVID-19 study and study participant characteristics.

Table 1. General information variables list

Variable	Description	Unit of reference	Coding	Filter	Notes	Version	Available	Derived
aid_baseline	CHRIS baseline Participant ID		<character>		Not available for family members of CHRIS baseline participants.		No	No
aid_invitation	CHRIS Covid-19 Participant ID of the invited participant		<character>		ID of the CHRIS baseline participant to whom the access token of the online questionnaire was associated.		No	No
cc_hhid	Identifier for individuals belonging to the same household		<character>		Derived from aid_invitation and from address information provided in the CHRIS Covid-19 registration and the CHRIS contact database.		Yes	Yes
cc_age	Age on February 1st, 2020, truncated to whole years	years	<integer>		February 1st, 2020 was set retrospectively as the starting date of the CHRIS Covid-19 study. The birth date was derived from the fiscal code.		Yes	Yes
cc_sex	Sex		1 Male 2 Female		Derived from the fiscal code.		Yes	No
cc_baseinfo	Baseline information available		<boolean>				No	Yes
cc_biochem	Biochemical traits available		<boolean>				No	Yes
cc_longinfo	Longitudinal information available		<boolean>				No	Yes
cc_vaccine	Vaccination status available		<boolean>				No	Yes
cc_rssele	Inclusion in random sample selection		<boolean>		Overall, 1812 CHRIS baseline participants were included in the initial random sample selection for a target of 1450 participants with 25% oversampling.		Yes	Yes
cc_rsexam	Participation at random sample examination		<boolean>				Yes	Yes

cc_esw	Expansion sampling weight		<float>	cc_rsexam=1			Yes	No
cc_rsw	Relative sampling weight		<float>	cc_rsexam=1			Yes	No

5. Advices for the analysis

Sex and age should be included in the analysis whenever appropriate.

The household identifier can be used to group participants as household members.

For analyses of the random sample, the use of the sample weights should be considered to improve the generalizability of the results.

6. References

Pattaro C, Barbieri G, Foco L, Weichenberger CX, Biasiotto R, De Grandi A, Fuchsberger C, Egger C, Amon VSC, Hicks AA, Mian M, Mahlknecht A, Lombardo S, Meier H, Weiss H, Rainer R, Dejaco C, Weiss G, Lavezzo E, Crisanti A, Pizzato M, Domingues FS, Mascalzoni D, Gögele M, Melotti R, Pramstaller PP. Prospective epidemiological, molecular, and genetic characterization of a novel coronavirus disease in the Val Venosta/Vinschgau: the CHRIS COVID-19 study protocol. Pathog Glob Health. 2022 Mar;116(2):128-136. doi: 10.1080/20477724.2021.1978225. PubMed PMID: 34637685