CHRIS Study

Participation information

Version 1.1

24th April 2024

Authors: LB, MG

1. Introduction

This module stores information related to the participation to the CHRIS study, such as examination date, the workflow position, the availability of genotyped data after quality control check, and the participation to the various substudies.

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".

2. History version changes

The cleaning process resulted in the following variables added:

variables added: x0_examm, x0_examy, x0_micros, aid_micros, barcode_micros, x0_nafld, aid_nafld, x0_familyla, x0_paregen, x0_geno, x0_wes.

3. Data cleaning

1. The main CHRIS dataset was loaded.
2. The notes variables related to the examination, namely x0_note, to the interview, x0_noteint, and to the self-administered questionnaire, x0_noteself, were translated and categorized into fewer groups when the content was similar.
3. The rows were sorted by order of appointment date.
4. The variable appointment, in string format, was transformed into date format yyyy-mm-dd and saved into the variable x0_examd.
5. From the variable x0_examd, the month and year of the examination variables were created: x0_examm and x0_examy.
6. The variable workflow was renamed as x0_workf.
7. The variables with notes describing deviations from protocol in specific modules and substudies were translated, categorized and grouped if with the same meaning.
8. The dataset with the MICROS study participation information was merged into the CHRIS dataset. The participants of MICROS only and not of CHRIS were removed. The resulting variable describing which rows had been matched, called _merge, had its value recoded into 0 (not matched) and 1 (matched), and then renamed as x0_micros. The aid and barcode variable of the MICROS study, microsaid and barcode, were renamed as aid_micros and barcode_micros.
9. The dataset with the CHRIS-NAFLD study participation information was merged into the CHRIS dataset. The participants of CHRIS-NAFLD only and not of CHRIS were removed. The resulting variable describing which rows had been matched, called _merge, had its value recoded into 0 (not matched) and 1 (matched), and then renamed as x0_nafld. The aid_nafld was transformed into a string variable.
10. The dataset with the Family LA study participation information was merged into the CHRIS dataset. The participants of Family LA only and not of CHRIS were removed. The resulting

variable describing which rows had been matched, called _merge, had its value recoded into 0 (not matched) and 1 (matched), and then renamed as x0_familyla.

11. The dataset with the CHRIS-Parkin RbG (PAREGEN) study participation information was merged into the CHRIS dataset. The participants of PAREGEN only and not of CHRIS were removed. The resulting variable describing which rows had been matched, called _merge, had its value recoded into 0 (not matched) and 1 (matched), and then renamed as x0_paregen.
12. The dataset with the genotype data availability in CHRIS, that already passed quality control screening, was merged into the CHRIS dataset. The participants with genotype data which passed the quality control who were not present in the CHRIS main dataset were removed. The resulting variable describing which rows had been matched, called _merge, had its value recoded into 0 (not matched) and 1 (matched), and then renamed as x0_geno.
13. The dataset with the whole exome sequence (WES) data availability in CHRIS, that already underwent quality control screening, was merged into the CHRIS dataset. The participants with WES data which passed the quality control who were not present in the CHRIS main dataset were removed (none in this case). The resulting variable describing which rows had been matched, called _merge, had its value recoded into 0 (not matched) and 1 (matched), and then renamed as x0_wes.
14. The baseline dataset was saved.

4. Advices for the analysis

Before starting with their analysis, the analyst should look at the note variables, to check if any missingness is due to the test not being completed or a lack of consent on that study module.

5. References

Pattaro C, Marroni F, Riegler A, et *al*. The genetic study of three population microisolates in South Tyrol (MICROS): study design and epidemiological perspectives. BMC Med Genet. 2007; 8, 29. DOI: 10.1186/1471-2350-8-29

Pattaro C, Gögele M, Mascalzoni D, 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

Klein C, Pramstaller PP, Kis B, et *al*. Parkin deletions in a family with adult-onset, tremor-dominant parkinsonism: expanding the phenotype. Ann Neurol. 2000 Jul;48(1):65-71. DOI: 10.1002/1531-8249(200007)48:1<65::AID-ANA10>3.0.CO;2-L

Mascalzoni D, Biasiotto R, Borsche M, *et al.* Balancing scientific interests and the rights of participants in designing a recall by genotype study. *Eur J Hum Genet* 2021;29; 1146–1157. DOI: 10.1038/s41431-021-00860-7

Motta BM, Grander C, Gögele M, et *al*. Microbiota, type 2 diabetes and non-alcoholic fatty liver disease: protocol of an observational study. J Transl Med. 2019 Dec 4;17(1):408. DOI: 10.1186/s12967-019-02130-z