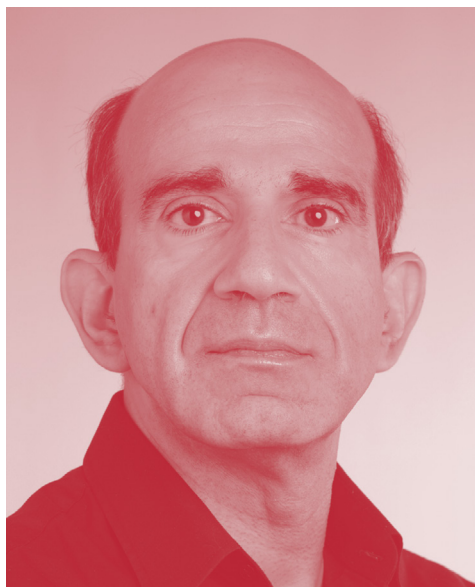# CASE STUDY

colaus
psycolaus

HOW THE COLAUS|PSYCOLAUS STUDY SHARES DATA WHILE ENSURING PRIVACY

Author: **Pedro Marques-Vidal**

Affiliations: Lausanne University Hospital (CHUV), Department of Medicine, Internal Medicine and University of Lausanne (UNIL)

While data sharing is important in an international research setting, ensuring the privacy of participants is just as important. For several years, the CoLaus|PsyCoLaus study has been sharing data with national and international research teams while ensuring that no breach of privacy regarding the participants occurs. A series of procedures have been put in place, including checking the research protocol, data encryption, and legally binding agreements. The original data are also distributed between sites and informatics systems, thus making re-identification of the participants difficult. These procedures allow a high level of security, and the participants are guaranteed that no data leakage will occur.

The CoLaus|PsyCoLaus study is an ongoing, prospective, population-based, cohort study investigating the relationships between cardiovascular and psychiatric diseases. Participants sign consent forms prior to the study, which are stored indefinitely in a secure location. Information collected for the study includes socio-economic, family and personal history of disease, medicines, lifestyle, clinical, biological, metabolomic, and genetic data. Over 1,000 variables and 7 terabytes of raw and processed data have been collected so far. One major issue is ensuring that this data is safely stored and shared without enabling the identification of the participants. To achieve this double objective, several procedures have been implemented.

1. DATA DISTRIBUTION

CoLaus|PsyCoLaus data is split into three geographically and informatically different sites (phenotypic, psychiatric, and genetic) that are under the responsibility of three different data managers who do not have direct access to each other's databases. Access to the databases is limited to the principal investigators, and passwords to enter the system are replaced regularly. Automatic backups are conducted regularly.

2. SELECTIVE COLLABORATION

Each research group that would like to use data from the CoLaus|PsyCoLaus team has to fill out a research protocol, which is evaluated by the study's scientific committee. There are several restrictions regarding data sharing. For instance, full genome data cannot be shared (but a limited number of genotypes can) and the number of variables requested must be justified. If the scientific committee finds a data request to be excessive, it can either reject the project or limit the number of variables provided. If a research group requesting data is located in a country whose legislation regarding data privacy is less stringent than Switzerland's, no data is provided. Similarly, data that could identify an individual (i.e. birthdate or geolocation) is either deleted (birthdate) or blurred (geolocation) before being sent.

3. LEGALLY BINDING DATA TRANSFER AGREEMENTS

If a research protocol is accepted by the scientific committee and the research group is outside the Lausanne University Hospital (**CHUV**) or the University of Lausanne (**UNIL**), a data transfer agreement (**DTA**) has to be signed by both parties. The legal office at CHUV has created a generic DTA template, which can be modified to suit both parties. The DTA states, among other things, that no individual participant data will be shared by the requesting research group, including in the publication of the results (see below). Each approved research protocol is given a number, and the protocols are stored in a dedicated folder within a server with limited access. The protocol title, contact information of the principal investigator, date of acceptance, duration of the research, and the study status (abandoned, research ongoing, publication, etc.) is entered in a registry that contains all research protocols approved by the CoLaus|PsyCoLaus scientific committee (over 300 as of March 2022). This registry makes it possible to contact research groups for an update on the status of their research.

4. RECORDED DATA EXTRACTION

After signing the DTA, data can be extracted. A statistical script code is written that indicates all the source databases used and all the variables extracted (or generated specifically for the research protocol). This code and the corresponding data are kept indefinitely in a specific folder for future checking. If a database is updated, all previous versions are kept.

5. ENCRYPTED DATA TRANSFER

If no secure email system is available, the data to be sent to an external research group is zipped and encrypted. The encrypted data is sent via email, and the password is sent via another channel, most frequently via SMS to the principal investigator. In some cases, instead of individual participant data, metadata is provided, such as frequencies, averages, and number of participants fulfilling a given condition.

6. DATA SHARING POLICY FOR PUBLICATIONS

There is an increasing number of journals that request the analysis database to be shared as a condition for publication. After consulting the cantonal ethics committee, it was concluded that such types of sharing would be a violation of the Swiss legislation that aims to protect the personal rights of participants. Hence, journals that explicitly request individual participant data are excluded from the publication strategy. A generic statement indicating that no individual participant data can be shared has been written and is copied and pasted in all papers submitted for publication. In any case, journal guidelines are subordinate to legislation, and it is the researcher, not the journal, who is legally responsible if a breach of privacy occurs.

CONCLUSION

When sharing research data, it is necessary to find a subtle balance between openness and participants' rights to privacy. It is imperative to implement procedures that ensure such a balance.