Clinical research data sharing in Switzerland in a nutshell

Sarah R Haile, Malwina Kowalska, Hester van de Wiel, Markus Golder, Leonhard Held, Ulrike Held

doi: 10.5281/zenodo.13860399

What is data sharing?

In recent years, funding agencies and journals increasingly require either data from clinical research projects to be publicly available, or clear instructions on how to access such data. Open data can be useful when conducting a systematic review, or generating new hypotheses leading to new research projects, but may also be crucial in terms of open science to verify the results of clinical research projects (Bauchner et al., 2016). Typical questions on data sharing are listed below, and answers to these questions are given in this primer.

1. A researcher publishes a paper in [a well-known journal] that requires data used in the analysis to be published alongside the manuscript as supplementary material, so that results, figures and tables are reproducible. What form should the data be in for the submission? What should be written in the data availability statement?

2. A team of researchers wants to perform a systematic review and meta-analysis using individual patient data. They find a relevant manuscript from a few years ago, and write to the corresponding author to get the data. How should the corresponding author respond to the request?

3. While reading a publication, a researcher realizes that a subgroup analysis by disease severity not detailed in the original analysis would help to plan a new study. They send the corresponding author a request for the data. Are the data still available?

4. A researcher would like to verify the results in a publication by reanalyzing the original data using their own code. They write to the corresponding author, but will the data be shared?

5. The funding agency requires a data management plan as a part of the funding application. Is future data sharing part of the data management plan?

Much has been written about data sharing (also known as open data) in the literature (Gahl et al., 2021), but it is not always clear what can be shared and how it should be shared. Here we provide a short overview of how health-related data from clinical research projects in Switzerland can be shared. We focus on how to share tabular data for specific studies, that is, what can be stored in spreadsheets.
We explicitly do not refer to ways of sharing genetic data or medical images, because they require additional measures to guarantee anonymization. Initiatives such as the Swiss Personalized Health Network, have focused more on general sharing of clinical research data nationwide, including development of infrastructure in hospitals to enable more widespread data accessibility. A broader discussion of such general data sharing is beyond the scope of this document.

Why share clinical data?

1. Funding agencies often have clear policies related to open research data. The Swiss National Science Foundation (SNSF) Open Research Data policy, for example, expects its funded researchers to a) store research data, b) share it with other researchers, and c) deposit it onto public repositories. Making data available is very often part of the data management plan.

2. Journals also increasingly request sharing of data to accompany a publication, and have added data availability statements to the list of other declarations (e.g. from Taylor & Francis, Springer, Wiley, BMJ or PLOS).

3. Conducting systematic reviews or generating new hypotheses benefits from open data, as it provides accessible and comprehensive datasets for analysis.

4. Verification of the results (reproducible science) is possible only if data is shared (Bauchner et al., 2016).

"Upon reasonable request"- is it enough?

It has become fairly common to see that data will be available "upon reasonable request." Such statements, however, are no guarantee that data will actually be available (Savage and Vickers, 2009; Vickers, 2006). Sharing individual subject data already at the time of publication is the best practice. It is a simple way to ensure that data will be available to future researchers, even if members of the original study team are no longer at the institution where the research was conducted.

What is allowed in Switzerland?

The legal basis for sharing of clinical research data in Switzerland is described in three different laws.

HRA Federal Act on Research involving Human Beings (Human Research Act), or Humanforschungsgesetz (HFG, SR 810.30): applicable to all research related to human diseases, with (human) health-related personal data.

FADP Federal Act on Data Protection, or Datenschutzgesetz (DSG, SR 235.1): relevant for federal institutions like ETH Zurich or companies under private law.

IDG Information and Data Protection Act of the Canton of Zurich (LS 170.4) : relevant for public institutions of the Canton of Zurich, such as the University of Zurich or the University Hospital Zurich. The IDG is currently being revised and will probably be aligned with the FADP in some respects.

The HRA describes conditions for further use of biological material or genetic data in Article 32, and of "non-genetic health-related personal data" in Article 33. Further use of either type of data, in original or pseudonymized form is permitted if participants have given their informed consent to do so. Pseudonymized means that subject IDs have been replaced with a code or alternative ID, which enables later query of participant data to a designated individual who has access to the list matching original identifiers to pseudonymized identifiers[1].
In contrast, anonymized data are defined by the HRA as data which cannot be traced back to a specific person, at least not without disproportionate effort (Article 3). In particular, the Health Research Ordinance (HRO, SR 810.301) notes that name, date of birth, address and unique identification numbers must be deleted or recoded for data to be considered anonymous (Article 25). Anonymization should be performed using current technology, and should be documented, including a description of remaining risk of reidentification (HRO Art. 25 and 26[2]). A close reading of the HRA implies that consent is not required to anonymize non-genetic health-related data (Article 32 has line 3 related to anonymization, while Article 33 does not). Even if informed consent for further use has not been explicitly given, Article 34 permits further use if the data have been anonymized.
Therefore, according to the HRA, anonymized data may be shared, even if study participants have not given their consent to sharing of their data. Pseudonymized data may only be shared if participants

[1] The terminology used for pseudonymization varies in the HRA according to language, in that German uses "verschlüsselt", French "codé", and Italian "codificato", and therefore the English translation is "coded".
[2] Update to the HRO as of November 1, 2024.

have consented to it. The principle investigator (PI) of a study is responsible for ensuring that sharing of study data conforms to the HRA.
For clinical trials, the study sponsor, e.g. University Hospital Zurich, is responsible for retaining data 20 years beyond the end of study (Clinical Trials Ordinance (ClinO, SR 810.305), Art. 45), regardless of whether it is shared with other researchers. For other studies, the study team should keep the data for at least 10 years beyond study end (HRO, Art. 23a).
Sensitive personal data, as defined by the FADP (Article 5), includes data related to health, genetic data, and biometric data uniquely identifying an individual. Both the FADP (§31, §39) and IDG-ZH (§9, §18) state that sensitive data should be anonymized as soon as possible, and only passed to third parties or published in such a way that individual subjects are not identifiable. The FADP does not apply to anonymized data if re-identification by a third party is impossible (Regulatory Affairs Platform of the SCTO, 2023).

What about animal patients' clinical research data?

In clinical research involving animal patients, it is important to note that animal health data is not personal data within the meaning of the Data Protection Act and does not fall within the scope of the HRA. When handling such data, it is crucial to remove any identifying information related to the animal's owner. Additionally, it is good practice to remove the animal's hospital number as it serves as a unique identifier. Efforts to anonymize data must consider that certain combinations of factors, such as rare breeds and region, might inadvertently lead to the identification of the owner. Therefore, obtaining informed consent (see below) is always advisable to ensure ethical and legal compliance.

How to share clinical data?

Informed Consent

Even if researchers do not plan to share anything but anonymized clinical research data, having informed consent on further use of the data ensures that sharing the data remains acceptable, even if the data are not as anonymous as intended. The HRO (Chapter 3) outlines what information study participants should receive. For example, Article 32 lists 4 points on which subjects should be informed in the case of "non-genetic health-related personal data in coded form." Accordingly, swissethics, the Swiss Association of Research Ethics Committees, has formulated a general consent for further use of coded personal data.

Anonymization

Data needs to be anonymized before it can be shared. Consent to data sharing by individual patients is not required if data is made fully anonymous. The principle investigator is responsible that the key linking personal data in the anonymized dataset to personal identifiers is not available. Anonymization includes, but is not restricted to,

- aggregating data,
- removing variables,
- reduction of precision in certain variables, e.g., grouping values or adding noise,
- restricting high and low values of a continuous variables to avoid identifiability of individuals.

There are various tools to anonymize data, such as amnesia, or the R package sdcMicro from the Swiss Data Anonymization Competence Center (SwissAnon). Regarding the question about the "strength" of anonymization, legal requirements state that data need to be anonymous as of current technical / digital possibilities. They need not be made anonymous regarding future potential developments in the direction of de-anonymization (as defined in HRO Articles 25 and 26).

FAIR principles

The FAIR principles have been proposed in 2016 with the intention to guide and improve Findability, Accessibility, Interoperability, and Reuse of digital assets, providing best practices for sharing data and metadata. A general structure for data preparation in a spreadsheet has been proposed by Broman and Woo (2018).

Repositories

Repositories where data can be made available include, but are not limited to, Zenodo, dryad, and Vivli. Data can be made fully available for download, or can be maintained under restricted access, possibly with an embargo period. In either case, metadata can be made available to ensure that data are findable. The SNSF has established two criteria as prerequisite for data sharing (following FAIR principles, non-commercial), and has made a list of repositories fulfilling these requirements.

License

The licenses state the conditions under which others may reuse your data. The Creative Commons CC-BY-NC-ND license is most restrictive, enables reuse in unadapted form only, solely for noncommercial purposes, and only so long as attribution is given to the author(s). The CC-BY license is less restrictive, requiring only attribution to the original authors, and is required by a few funding agencies, particularly in Europe. The SNSF only requires that it be possible for the repository to assign a license to the data, but does not require a specific license to be used. More information on Creative Commons licenses can be found on their website and the SNSF also provides an overview.

Multinational Studies

A special case occurs if data was collected not only in Switzerland but also in other countries, as is typical in multi-center studies. In this case, different legal bases may apply, such as the European Union's General Data Protection Regulation (GDPR), as for example described by Bentzen et al. (2021).

More information

Simple rules for data sharing in clinical trials have been summarized in Pellen et al. (2023), with a broader discussion of the issues surrounding data sharing in Ohmann et al. (2017). Key steps in data sharing, along with many useful resources, have been described, for example, in the CRS Primer Sharing Clinical Trial Data (van de Wiel et al., 2024) and by Open Research Europe. In the Swiss context, barriers to sharing of health data have also been examined (Ormond et al., 2024) in a qualitative study. The Swiss Clinical Trials Organization has published extensive recommendations on Sharing of Data from Clinical Research Projects (Gahl et al., 2021). The Swiss Open Research Data Strategy Council supports researchers in adopting open research data practices, and the European project Share CTD researches various aspects of data sharing (Mansmann et al., 2023).

References

Bauchner, H., Golub, R. M., and Fontanarosa, P. B. (2016). Data Sharing: An Ethical and Scientific Imperative. *JAMA*, 315(12):1238–1240. doi:10.1001/jama.2016.2420.

Bentzen, H. B., Castro, R., Fears, R., Griffin, G., ter Meulen, V., and Usrin, G. (2021). Remove obstacles to sharing health data with researchers outside of the European Union. *Nature Medicine*, 27:1329–1333. doi:10.1038/s41591-021-01460-0.

Broman, K. W. and Woo, K. H. (2018). Data organization in spreadsheets. *The American Statistician*, 72(1):2–10. doi:10.1080/00031305.2017.1375989.

Gahl, B., Haynes, A. G., Sluka, C., Dupuis-Lozeron, E., Joerger, F., Schur, R., Christen, A., and Trelle, S. (2021). Sharing of data from clinical research projects: Guidance from the SCTO's clinical trial unit network. doi:10.54920/SCTO.2021.02.

Mansmann, U., Locher, C., Prasser, F., Weissgerber, T., Sax, U., Posch, M., Decullier, E., Cristea, I. A., Debray, T. P. A., Held, L., Moher, D., Ioannidis, J. P. A., Ross, J. S., Ohmann, C., and Naudet, F. (2023). Implementing clinical trial data sharing requires training a new generation of biomedical researchers. *Nature Medicine*, 29:298–301. doi:10.1038/s41591-022-02080-y.

Ohmann, C., Banzi, R., Canham, S., Battaglia, S., Matei, M., Ariyo, C., Becnel, L., Bierer, B., Bowers, S., Clivio, L., Dias, M., Druml, C., Faure, H., Fenner, M., Galvez, J., Ghersi, D., Gluud, C., Groves, T., Houston, P., Karam, G., Kalra, D., Knowles, R. L., Krleža-Jerić, K., Kubiak, C., Kuchinke, W., Kush, R., Lukkarinen, A., Marques, P. S., Newbigging, A., O'Callaghan, J., Ravaud, P., Schlünder, I., Shanahan, D., Sitter, H., Spalding, D., Tudur-Smith, C., van Reusel, P., van Veen, E.-B., Visser, G. R., Wilson, J., and Demotes-Mainard, J. (2017). Sharing and reuse of individual participant data from clinical trials: principles and recommendations. *BMJ Open*, 7(12). doi:10.1136/bmjopen-2017-018647.

Ormond, K. E., Bavamian, S., Becherer, C., Currat, C., Joerger, F., Geiger, T. R., Hiendlmeyer, E., Maurer, J., Staub, T., and Vayena, E. (2024). What are the bottlenecks to health data sharing in Switzerland? an interview study. *Swiss Medical Weekly*, 154(1):3538. doi:10.57187/s.3538.

Pellen, C., Le Louarn, A., Spurrier-Bernard, G., Decullier, E., Chrétien, J.-M., Rosenthal, E., Le Goff, G., Moher, D., Ioannidis, J. P. A., and Naudet, F. (2023). Ten (not so) simple rules for clinical trial data-sharing. *PLOS Computational Biology*, 19(3):1–10. doi:10.1371/journal.pcbi.1010879.

Regulatory Affairs Platform of the SCTO (2023). The new federal act on data protection (nFADP): Guidance for clinical trials. doi:10.54920/SCTO.2023.02.

Savage, C. J. and Vickers, A. J. (2009). Empirical study of data sharing by authors publishing in PLoS journals. *PLOS ONE*, 4(9):e7078. doi:10.1371/journal.pone.0007078.

van de Wiel, H., Burgwinkel, C., and Held, L. (2024). Sharing clinical trial data. *CRS Primer*. doi:10.5281/zenodo.13860164.

Vickers, A. J. (2006). Whose data set is it anyway? Sharing raw data from randomized trials. *Trials*, 7(15). doi:10.1186/1745-6215-7-15.

Version from September 30, 2024.

Edited by Sarah Haile (0000-0002-4704-6570) and Ulrike Held (0000-0003-3105-5840).