

Implementing clinical trial data sharing requires training a new generation of biomedical researchers

Ulrich Mansmann, Clara Locher, Fabian Prasser, Tracey Weissgerber, Ulrich Sax, Martin Posch, Evelyne Decullier, Ioana A. A. Cristea, Thomas. P. A. Debray, Leonhard Held, et al.

▶ **To cite this version:**

Ulrich Mansmann, Clara Locher, Fabian Prasser, Tracey Weissgerber, Ulrich Sax, et al.. Implementing clinical trial data sharing requires training a new generation of biomedical researchers. Nature Medicine, 2023, 29 (2), pp.298-301. 10.1038/s41591-022-02080-y . hal-04010046

HAL Id: hal-04010046
https://hal.science/hal-04010046

Submitted on 31 Mar 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Implementing clinical trial data sharing requires training a new generation of biomedical researchers

Version: 11.09.2022 11:00

NMED-C120847B

Ulrich Mansmann[1] (ORCID: 0000-0002-9955-8906)
Clara Locher[2] (ORCID: 0000-0002-8212-4351)
Fabian Prasser[3] (ORCID: 0000-0003-3172-3095)
Tracey Weissgerber[3] (ORCID: 0000-0002-7490-2600)
Ulrich Sax[4] (ORCID: 0000-0002-8188-3495)
Martin Posch[5] (ORCID: 0000-0001-8499-8573)
Evelyne Decullier[6] (ORCID: 0000-0002-2412-4767)
Ioana A. Cristea[7] (ORCID: 0000-0002-9854-7076)
Thomas P.A. Debray[8,9] (ORCID: 0000-0002-1790-2719)
Leonhard Held[10] (ORCID: 0000-0002-8686-5325)
David Moher[11, 12] (ORCID: 0000-0003-2434-4206)
John P.A. Ioannidis[13] (ORCID: 0000-0003-3118-6859)
Joseph S. Ross[14, 15] (ORCID: 0000-0002-9218-3320)
Christian Ohmann[16] (ORCID: 0000-0002-5919-1003)
Florian Naudet[2,17] (ORCID 0000-0003-3760-3801)

Keywords: Clinical trials, individual patient data, open science, data sharing, training

Accepted manuscript / Final version

1: Ludwig-Maximilians University Munich, Institute for Medical Information Processing, Biometry, and Epidemiology, München, Germany.

2: Université Rennes, CHU Rennes, Inserm, Centre d'investigation clinique de Rennes (CIC1414), service de pharmacologie clinique, Institut de recherche en santé, environnement et travail (Irset), UMR S 1085, EHESP, 35000, Rennes, France

3: Berlin Institute of Health at Charité - Universitätsmedizin Berlin, Germany

4: Department of Medical Informatics, University Medical Center, Göttingen, Germany.

5: Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Austria.

6: Hospices Civils de Lyon, Pôle Santé Publique, Unité de Recherche Clinique, Lyon, France.

7: Department of Brain and Behavioral Sciences, University of Pavia, Pavia, Italy.

8: Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands.

9: Smart Data Analysis and Statistics B.V., Utrecht, The Netherlands.

10: Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Switzerland.

11: Center for Journalology, Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Canada.

12: School of Epidemiology and Public Health, University of Ottawa, Ottawa, Canada.

13: Departments of Medicine, of Epidemiology and Population Health, of Biomedical Data Science, and of Statistics, and Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, USA

14: Section of General Medicine, Department of Internal Medicine, Yale School of Medicine;

15: Department of Health Policy and Management, Yale School of Public Health; Center for Outcomes Research and Evaluation, Yale New Haven Health System, all New Haven, CT, USA.

16: European Clinical Research Infrastructure Network (ECRIN), Düsseldorf, Germany.

17: Institut Universitaire de France (IUF)

Corresponding author: Ulrich Mansmann

Accepted manuscript / Final version

Data sharing enhances the value of medical research and builds trust in clinical trials, but more biomedical researchers need to be trained in these approaches, which include meta-research, data science and ethical, legal and social issues.

Clinical trials form foundational evidence to inform contemporary medical decision making. They provide evidence widely used by regulatory bodies, health technology assessment agencies, and are considered the gold standard for assessing treatment effects. The value and trustworthiness of medical research may be enhanced by sharing of patient-level clinical trial data together with the code on which analyses are based (1, 2) as well as other materials such as the protocol, case report form, and data dictionary.

Embedding clinical trial data sharing into such a broader framework offers the opportunity for external re-analysis, which enables conclusions to be re-examined, verified or, occasionally, corrected, thereby building trust. Data sharing also allows individual participant data (IPD) meta-analysis and other strategies that build upon previous data and code, such as secondary analyses and methodological work. Data sharing should accelerate discovery, reduce false discovery rates, and potentially discourage misconduct and research waste, as well as allowing more value to be drawn from the original research investment. Data sharing honours the generosity of clinical trial participants, because it maximises the utility of the data they provide (3), and is widely viewed as a positive feature by stakeholders involved in clinical trials, including trial participants (4).

Clinical trial data sharing experts are therefore urgently needed. A new generation of such experts can be nurtured by incorporating interdisciplinary methodological approaches to clinical trial data sharing into the curriculum of existing medical PhD and Clinical Scientist programs around the globe.

Limited sharing

Over the past decade, data sharing platforms have been launched to promote clinical trial data sharing, including Clinical Study Data Request (CSDR), the Yale University Open Data Access (YODA) Project, and Vivli. Guidelines have been developed to verify digital repository trustworthiness (5). Regulatory authorities such as the European Medicines Agency (EMA) and the US Food and Drug Administration launched initiatives on clinical trial data transparency (6). Although their implementation is a work in progress, EMA’s policy 0070 on “publication of clinical data for medicinal products for human use” includes plans to publish the IPD (7). The pharmaceutical industry also adopted policies for sharing IPD (8). Many public funders (9) and medical journals (10) support, and sometimes even mandate, data sharing, including Wellcome

Accepted manuscript / Final version

Trust, the British Medical Journal and Public Library of Science. Academic and commercial communities are developing best practices on how to organize and perform data sharing.

Despite these positive developments, the practical implementation of clinical trial data sharing is disappointing. Over the 2017-2019 period, about 16% of pivotal trials submitted to the EMA provided IPD for re-analysis (11). Data sharing policies are poorly implemented by journals and rarely adhered to. In surgical journals no change was observed in data sharing before (2/65 trials) and after (2/65 trials) the International Committee of Medical Journal Editors adopted a data sharing policy (12).

There is a lack of adequate incentives to fully implement clinical trial data sharing. Scientific productivity is currently favoured over transparency in promotion and tenure criteria in biomedical sciences faculties in academic medicine (13). Some trialists may be sceptical about data sharing; with some re-users labelled as “research parasites” (14). Researchers and data-providers may also lack the adequate support and/or financial resources and face technical barriers, such as a lack of secure infrastructure to handle requests, or prepare and share their data sets. There can be technical hurdles, such as additional effort that is often needed to harmonize variables across data sets. Trialists may also face regulatory difficulties in sharing their data, as health data are sensitive and require adequate privacy protection. In addition, some poor-quality data requests may make the data generator reluctant to share the data. These challenges limit the impact of data sharing efforts, providing a compelling need to improve data sharing processes in clinical trials.

New professional activities for experts can help overcome cultural and practical barriers and improve clinical trial data sharing efforts. The next generation of biomedical researchers should be trained in their specific domains, and the entire lifecycle of clinical trial data sharing (**Figure 1**). The scarcity of such knowledgeable experts to-date presents a key bottleneck in accelerating data sharing efforts, as well as hindering the adequate use of shared data. Contemporary training approaches are usually focused on individual components of clinical trial data sharing, whereas an interdisciplinary approach is needed. Such training should focus on implementing clinical trial data sharing and assessing its impact, both by preparing and reusing the data.

A training curriculum

Organising data sharing activities both at the study level (requesting, preparing, sharing data and re-using the data) and at a global level (adopting and optimising data sharing policies) requires an inter- and transdisciplinary approach that includes clinical trials regulations, ethical, legal, and

Accepted manuscript / Final version

social issues, informatics, data science, biostatistics, and meta-research, as well as domain expertise across different medical fields.

Those who establish best practice for clinical trial data need training in principles, governance, skills, and operations. Such training should include best practices, measuring impact, and practical exercises for the use of shared data.

Best practice

A coherent training curriculum should encompass the many topics related to best practices in clinical trial data sharing. Best practice in preparing an existing data set for sharing requires standards including the FAIR (15, 16) and TRUST Data Principles (5), privacy protection of trial participants, efficient anonymization, reporting of studies, and governance processes to access data. Best practice in re-using a shared data set include the access and utilization of the data. This requires skills in handling data formats, managing deployment processes (such as data sharing platforms such as YODA or federated and distributed analyses that are needed when data cannot leave the host institution), implementing secondary studies such as IPD meta-analysis (17, 18), clinical trial planning and biostatistical methods development. Obviously, skills in open science and reproducible research practices are paramount as transparent reporting is needed to ensure that results can be properly interpreted and reproduced. Scientists with awareness of these topics should be ready to collaborate with the right experts, all of whom needed for a successful project set-up.

Impact

While best practices for clinical trial data sharing are coming progressively into effect, it is necessary to assess whether they achieve the intended impact. Therefore, the training curriculum should encompass the various aspects related with understanding and studying the impact of clinical trial data sharing. Training should enable those who perform clinical trials to assess how various stakeholders, including trial participants, understand and accept data sharing (4). Furthermore, researchers must understand the cost associated with data preparation and data curation and its implications.

At the more global level, impact studies should assess how clinical trial data sharing influences knowledge generation. This includes the study of indicators of transparency, openness, and reproducibility for clinical trials. Such activities require the use of automated screening tools (19) to identify studies relying on shared data and monitor their conduct, transparency and reporting.

Accepted manuscript / Final version

This part of the curriculum requires teaching skills on a wide range of meta-research approaches, such as umbrella reviews, surveys, observational studies, and simulations (19).

Datathons for data literacy

Biomedical researchers are usually trained with theoretical lectures and practical exercises. Although these activities help to develop familiarity with key data sharing concepts, more interactive methods such as datathons should be used to acquire practical skills. A datathon is an intense workshop that asks participants to utilise the data provided to develop and answer topic-driven questions or to develop innovative approaches to analyse the data.

During these events, participants gather to solve practical problems through the application of data science tools and techniques, working together in teams to generate insights and potential solutions. Participants in a datathon can attempt to make a novel clinical finding from a single RCT dataset, as for the Systolic Blood Pressure Intervention Trial (SPRINT) Challenge (20). Each team defines an interesting research question requiring data re-use and explores the question using the datasets. This secondary analysis of clinical trial data could investigate subgroup effects or rare safety events, or explore reproducibility issues (21), or implement an IPD meta-analysis (17).

Participants can benefit from the input of experts within the project as well as external advisors such as patients and clinicians. Beyond teaching skills, datathons may also teach values by promoting transparency, honesty and collaboration in an environment where ideas can be shared openly and implementation can become more efficient. Such multi-team analyses stands in contrast to the overly optimistic and unrepresentative results from a single statistical analysis (22).

Funding requirements

All organizations that conduct and report clinical trials need to train a new generation of scientists who are able to understand the challenges related to clinical trial data sharing and to implement open science practices that maximize their value, transparency and reproducibility. Funding will be needed, as will buy-in from stakeholders in curricular and strategic development, as well as management in academia and industry.

Typically, clinical trial data sharing represents a complex task with many parts which are not reflected in the traditional priorities for research funders, although funders prioritize open science. UNESCO has named open practices in science as one of their priorities (23), and the French Plan for Open Science has a specific working group dedicated to clinical trial data sharing. The

Accepted manuscript / Final version

University of Cambridge has a well-established data champions programme (24), training research volunteers who advise their peers on best practices related with research data management, including data sharing. Canada has funded 18 data champions programs (25), some of which focus on clinical research, including clinical trials. The technical university of Delft counts data champion activities towards career advancement (22).These initiatives align with DORA (26) and Hong Kong Principles (27), which aim to reform research assessment in academia by explicitly recognizing and rewarding research integrity and reproducible research practices.

Additional professional opportunities for clinical trial data sharing arise in universities, research institutions, and pharmaceutical companies, settings in which clinical trials are funded, conducted and reported (Box 1). (27).

Only through training a new generation of data sharing experts can the full potential of clinical trial data be realized for the advancement of medical research and the benefit of patients.

Accepted manuscript / Final version

1. Each clinical trial data sharing platform needs experts for a complex, thorough, and efficient review process. Experts will have their own specific field, such as medicine, biostatistics, interdisciplinary data science, intellectual property, privacy protection, ethical, legal, and social issues, but also need an overarching understanding of the disciplinary nature of this topic.
2. The peer review of research papers that come from clinical trial data sharing should reflect the interdisciplinary approach and complexity. Editors or editorial staff should boost clinical trial data sharing within their journals' activities. Requiring data and analysis scripts for reviews will initiate basic processes to share clinical trial data. Reproducibility editors can steer the verification of published results based on the data used.
3. Academic institutions, such as trial centres at academic hospitals, need well-trained staff, including data managers and data engineers. These staff can set up and contribute to clinical trial data sharing activities, especially for anonymization requirements.
4. Research funders, as well as research management, in academia, industry, governments, and regulatory bodies must be trained in the complex issues of clinical trial data sharing.
5. Future medical researchers, as well as data scientists in medicine, will need basic training on clinical trial data sharing so that they can conduct their own research using shared datasets.
6. Training in meta-research for academics, professional institutions and funders will maximise the impact of data sharing.

Box 1: Opportunities for experts in clinical trial data sharing

Accepted manuscript / Final version

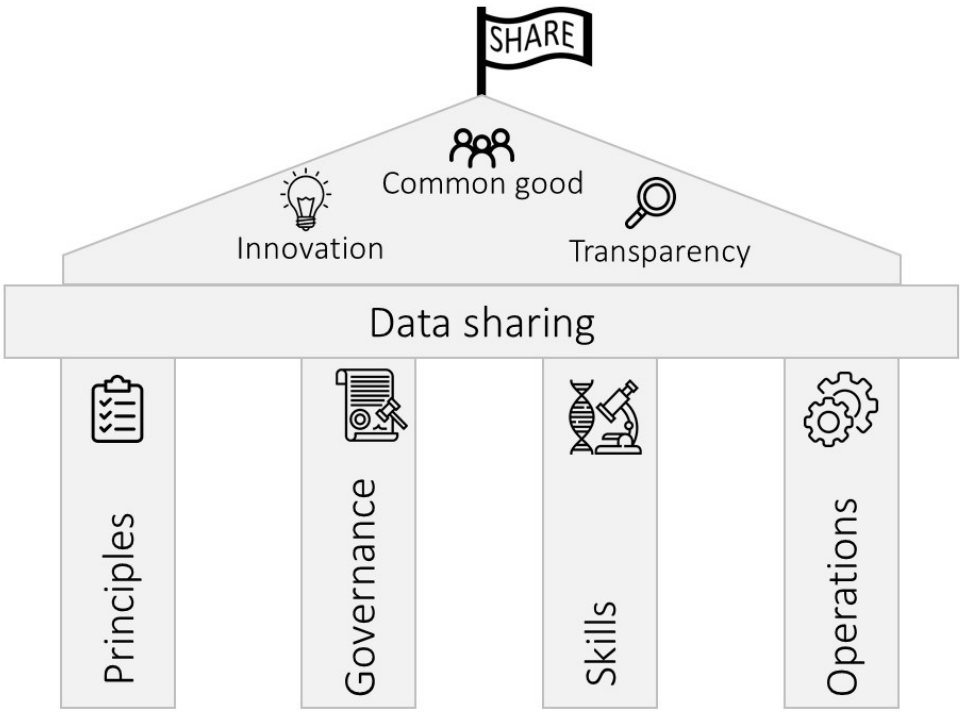


Figure 1: Elements of data sharing

Data sharing is built on principles, governance structures, skills, and operation infrastructure. It shapes scientific openness, transparency, and reproducibility as virtues of a scientific community which demonstrates good practice and supports change.

Accepted manuscript / Final version

References:

1. Koenig F et al. Biom J. 2015;57(1):8-26.
2. Localio AR et al. Ann Intern Med. 2018;168(11):828-9.
3. Taichman DB et al. PLoS Med. 2017;14(6):e1002315.
4. Mello MM et al. N Engl J Med. 2018;378(23):2202-11.
5. Lin D et al. Sci Data. 2020;7(1):144.
6. Egilman AC et al. J Law Med Ethics. 2021;49(3):456-85.
7. EMA 2019 https://www.ema.europa.eu/en/documents/other/european-medicines-agency-policy-publication-clinical-data-medicinal-products-human-use_en.pdf (accessed 06.09.2022)
8. Sudlow R et al. BMC Med Res Methodol. 2016;16 Suppl 1(Suppl 1):73.
9. Gaba JF et al. PLoS One. 2020;15(8):e0237464.
10. Naudet F et al. PLoS Med. 2021;18(10):e1003844.
11. Siebert M et al. BMC Med. 2022;20(1):177.
12. Bergeat D et al. JAMA Netw Open. 2022;5(6):e2215209.
13. Rice DB et al. BMJ. 2020;369:m2081.
14. Longo DL, Drazen JM. N Engl J Med. 2016;374(3):276-7.
15. Gouthamchand V et al. medRxiv. 2022:2021.07.23.21261032.
16. Wilkinson MD et al. Sci Data. 2016;3:160018.
17. Tierney JF et al. PLoS Med. 2015;12(7):e1001855.
18. Debray TP et al. PLoS Med. 2015;12(10):e1001886.
19. Weissgerber T et al. Nat Med. 2021;27(1):6-7.
20. Burns NS et al. N Engl J Med. 2017;376(23):2205-7.
21. Goodman SN et al. Sci Transl Med. 2016;8(341):341ps12.
22. Wagenmakers EJ et al. Nature. 2022;605(7910):423-5.
23. UNESCO 2021 https://en.unesco.org/science-sustainable-future/open-science/recommendation (accessed 06.09.2022)
24. University C, https://www.data.cam.ac.uk/intro-data-champions (accessed 06.09.2022)
25. Canada DRA 2022, https://alliancecan.ca/en/latest/news/announcing-2022-2023-data-champions (accessed 06.09.2022)
26. DORA, https://sfdora.org/read/ (accessed 96.09.2022).
27. Moher D et al. PLoS Biol. 2021;19(4):e3001140.

Accepted manuscript / Final version

Author contributions

All authors contributed to the concept that the manuscript presents. Ulrich Mansmann and Florian Naudet wrote the manuscript that was carefully and critically proofread by all authors. All authors have given their consent to the publication of the manuscript.

Accepted manuscript / Final version

Conflict of Interest:

Ulrich Mansmann[1] (ORCID: 0000-0002-9955-8906): No conflict of interest

Clara Locher[2] (ORCID: 0000-0002-8212-4351): No conflict of interest

Fabian Prasser[3] (ORCID: 0000-0003-3172-3095): No conflict of interest

Tracey Weissgerber[3] (ORCID: 0000-0002-7490-2600): No conflict of interest

Ulrich Sax[4] (ORCID: 0000-0002-8188-3495): No conflict of interest

Martin Posch[5] (ORCID: 0000-0001-8499-8573): No conflict of interest

Evelyne Decullier[6] (ORCID: 0000-0002-2412-4767): No conflict of interest

Ioana A. Cristea[7] (ORCID: 0000-0002-9854-7076): No conflict of interest

Thomas P.A. Debray[8,9] (ORCID: 0000-0002-1790-2719): Owns a SME that provides consulting services to life science industries. He declares that no support was received when contributing to this paper.

Leonhard Held[10] (ORCID: 0000-0002-8686-5325): No conflict of interest

David Moher[11, 12] (ORCID: 0000-0003-2434-4206): Receives funding from the Digital Research Alliance of Canada to develop a Data Champions program in Medicine.

John P.A. Ioannidis[13] (ORCID: 0000-0003-3118-6859): No conflict of interest

Joseph S. Ross[14, 15] (ORCID: 0000-0002-9218-3320): Dr. Ross currently receives research support through Yale University from Johnson and Johnson to develop methods of clinical trial data sharing.

Christian Ohmann[16] (ORCID: 0000-0002-5919-1003): No conflict of interest

Florian Naudet[2,17] (ORCID 0000-0003-3760-3801): No conflict of interest

Accepted manuscript / Final version