UNIVERSITÄT BERN

Clinical Data Sharing

Chancen & Herausforderungen

Sven Trelle
CTU Bern
University of Bern

14.06.2017, CTU Lecture

Outline

- Background and rationale
- Examples
- Types of data sharing
- Current landscape
- Support

Rationale

- Ethical obligation to participants
- Transparency
- Enable reproducibility/replication
  - Reproducing original analyses
  - Replicating the original study
- Enable additional scientific discoveries
  - Additional/new analyses
  - Meta-analysis

Early Breast Cancer Trialists' Collaborative Group (1985)

UNIVERSITÄT BERN

Outline

Summary
Introduction
Methods
Results
Discussion
Age
Nodal status
Addition of polychemotherapy to tamoxifen
Different polychemotherapy regimens
Conclusions
References

Show full outline

Figures (8)

Show all figures

THE LANCET

Volume 352, Issue 9132, 19 September 1998, Pages 930-942

Articles

Polychemotherapy for early breast cancer: an overview of the randomised trials *

Early Breast Cancer Trialists' Collaborative Group

Show more

https://doi.org/10.1016/S0140-6736(98)03301-7 Get rights and content

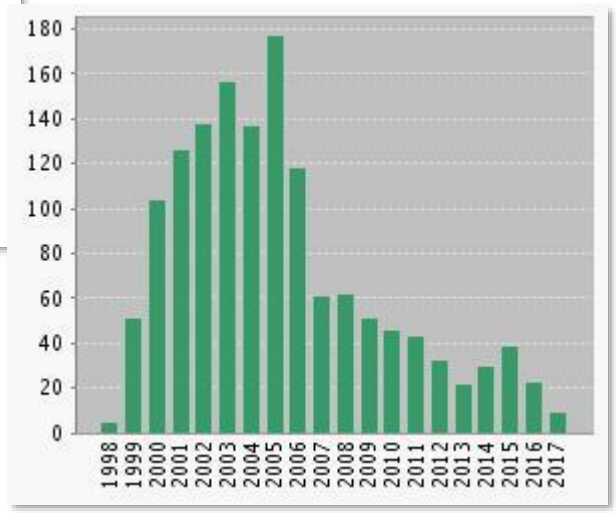
Summary

Background

There have been many randomised trials of adjuvant prolonged polychemotherapy among women with early breast cancer, and an updated overview of their results is presented.

Methods

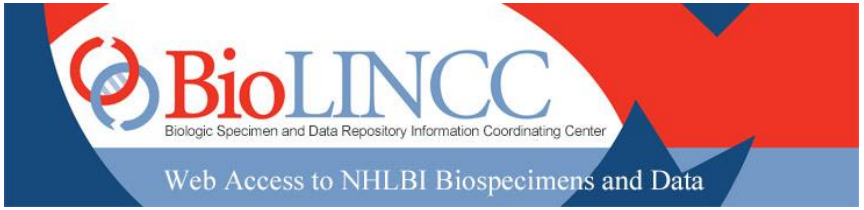
In 1995, information was sought on each woman in any randomised trial that began before 1990 and involved treatment groups that differed only with respect to the chemotherapy regimens that were being compared. Analyses involved about 18 000 women in 47 trials of prolonged polychemotherapy versus no chemotherapy, about 6000 in 11 trials of longer versus shorter polychemotherapy, and about 6000 in 11 trials of anthracycline-containing

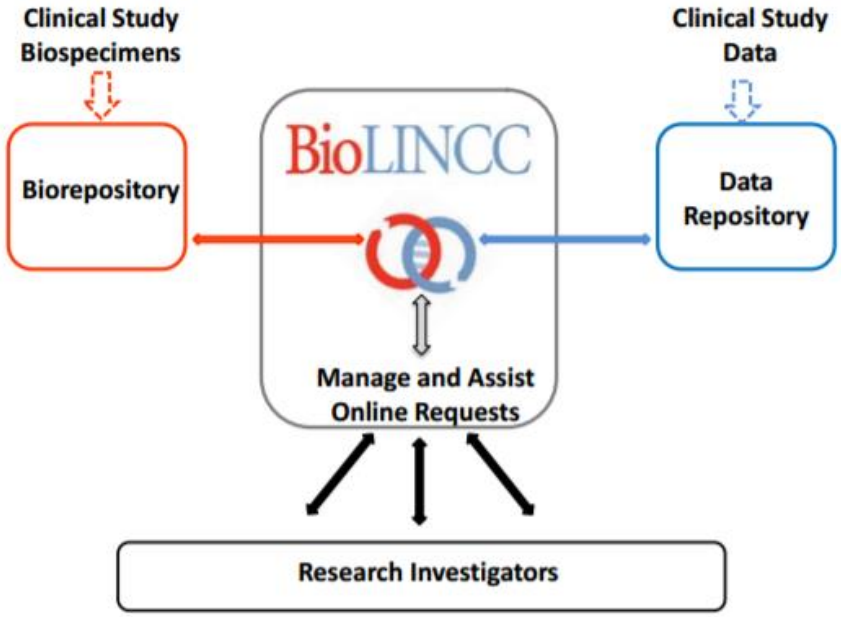


Early Breast Cancer Trialists' Collaborative Group 1998

National Heart, Lung, and Blood Institute (National Institutes of Health)

- Established 2008
- 169 studies (06/2017)





Types of data sharing

UNIVERSITÄT BERN

- Data?
  - Individual participant data, the study database
  - Meta information
    - Data dictionary
    - Case Report Forms
    - Study protocol
    - Patient information and informed consent form
    - Data Management Plan
    - Statistical Analysis Plan
    - …

Types of data sharing

UNIVERSITÄT BERN

- Freely available to everybody without any restrictions
  - Very rare if at all done in patient-oriented clinical research
  - E.g. US National Cancer Institute, Surveillance, Epidemiology, and End Results (SEER) Program (only username/password)
- Available on request with different levels of requirement
  - No review but agreement
    - E.g. Enroll-HD (www.enroll-hd.org) → "researcher employed by a recognized academic institution, company or nonprofit organization"; agreement to uphold privacy and obligation to publish
  - Review and agreement
    - E.g. BioLINCC
  - Review, agreement, and involvement of original researcher(s)
    - E.g. Swiss HIV Cohort Study

Types of data sharing

- From a person
- Via study website
  - Often large and long running cohort studies
  - E.g. Danish National Birth Cohort (http://www.ssi.dk/English/RandD/Research%20areas/Epidemiology/DNBC.aspx)

Danish National Birth Cohort

› Access to DNBC data
  › Conditions for access to data
  › How to apply for data
  › Data available
  › DNBC's formalities

Back to Epidemiology

How to apply for data

All researchers with a project that falls within the overall aim of the DNBC may apply for data.

Request Application Form

Please start by requesting the newest version of the DNBC application form. Send your request to: dnbc-research@ssi.dk.
Fill it in and return it together with a short protocol to: dnbc-research@ssi.dk.

Your application will be given a reference number and will be submitted to the DNBC Management and then to the DNBC Steering Committee. You can expect to hear from us again after 6-8 weeks.

If you are asking for data where individuals may be identified, you must seek permission with the Danish Data Protection Agency (Datatilsynet).

If you are applying for biological material you also need permission from the Committee on Biomedical Research Ethics (Videnskabsetisk komité).

Your dataset is available on our server

To enhance data security your dataset will be made available to you through a secure connection to our server at Aarhus University.
Please direct your inquiries to data managers Lone Fredslund (lf@ph.au.dk) or Inge Eisensee (ie@ph.au.dk).

The cost of data

The price for your dataset will depend on the time our data managers spend on assisting you and constructing your dataset. Usually, the price will be around DKK 10 - 15,000. The connection to the server including storage space costs DKK 5,000 a year.

Bedre sundhed i generationer

Contact

Danish National Birth Cohort
Inger Kristine Meder
Project Coordinator

Tel: + 45 3268 8121
dnbc-research@ssi.dk

Accessing data with Citrix

Below you will find the terms for accessing your DNBC-dataset via Citrix.

Access to data from the DNBC - general information

Types of data sharing

- From a person
- Via study website
  - Often large and long running cohort studies
  - E.g. Danish National Birth Cohort (http://www.ssi.dk/English/RandD/Research%20areas/Epidemiology/DNBC.aspx)
- Via data repository (Registry of Research Data Repositories; http://www.re3data.org)
  - E.g. Clinical trial data from selected pharmaceutical companies (CSDR; https://www.clinicalstudydatarequest.com/)
  - E.g. Yale University Open Data Access (YODA; http://yoda.yale.edu/)

FAIR: origin and legitimation

UNIVERSITÄT BERN

- FORCE11
  - «Movement« of researchers, librarians, archiving specialists, publisher, funding agencies
  - Currently >2’000 members
- Workshop in 2014 in Leiden, Netherlands
  - «Jointly Designing a Data Fairport»
  - → FAIR Guiding Principles

- Supported/cited
  - EU HORIZON 2020
  - Swiss National Science Foundation

What is special about FAIR?

UNIVERSITÄT BERN

- Many discipline-specific initiatives

- FAIR
  - Short but precise
  - Independent of any research discipline
  - Principles
  - Broadly applicable (not only for data itself)

Broadly applicable

UNIVERSITÄT BERN

- Data and metadata
- But also the research that led to the data
  - Algorithms
  - Tools, scripts
  - Processes and workflows
  - …
- All parts of the research process must be available/accessible
  - Reproducibility
  - Transparency
  - Sustainability
  - Further use

They are (not)

- Principles!
- No answers on
  - Implementation
  - Technology
  - Standards
  - Specifications
  - …

FAIR =

- Findable (auffindbar)
- Accessible (verfügbar)
- Interoperable (kompatibel)
- Re-usable (wiederverwendbar)

Findable

- F1. (Meta)daten haben eine eindeutige und ewig bestehende Kennung
- F2. Daten sind mit reichhaltigen Metadaten beschrieben
- F3. (Meta)daten sind registriert oder indexiert in einer auffindbaren/suchbaren Quelle
- F4. Metadaten beschreiben die Kennung

Accessible

- A1. (Meta)daten sind abrufbar durch ihre Kennung mittels standardisiertem Protokoll
  - A1.1. Das Protokoll ist offen, frei und universell einführbar
  - A1.2. Das Protokoll erlaubt einen Authentifizierungs- und Authorisierungsprozess falls erforderlich
- A2. Metadaten sind zugänglich/verfügbar auch wenn die eigentlichen Daten nicht (mehr) verfügbar/vorhanden sind z.B. aus Datenschutzgründen

Interoperable

- I1. (Meta)daten verwenden eine zugängliche, gemeinsame und breit anwendbare Sprache für Beschreibungen («knowledge representation»)
- I2. (Meta)daten benutzen ein Vokabular, dass den FAIR Prinzipien folgt
- I3. (Meta)daten referenzieren andere (Meta)daten

Re-usable

UNIVERSITÄT BERN

- R1. Meta(daten) haben eine Vielzahl an korrekten und relevanten Attributen
  - R1.1. (Meta)daten werden mit einer eindeutigen und zugänglichen Nutzungslizenz freigegeben
  - R1.2. (Meta)daten haben eine Herkunftsbezeichnung (Ursprung ist klar)
  - R1.3. (Meta)daten erfüllen die fachspezifischen Standards

More Information

UNIVERSITÄT BERN

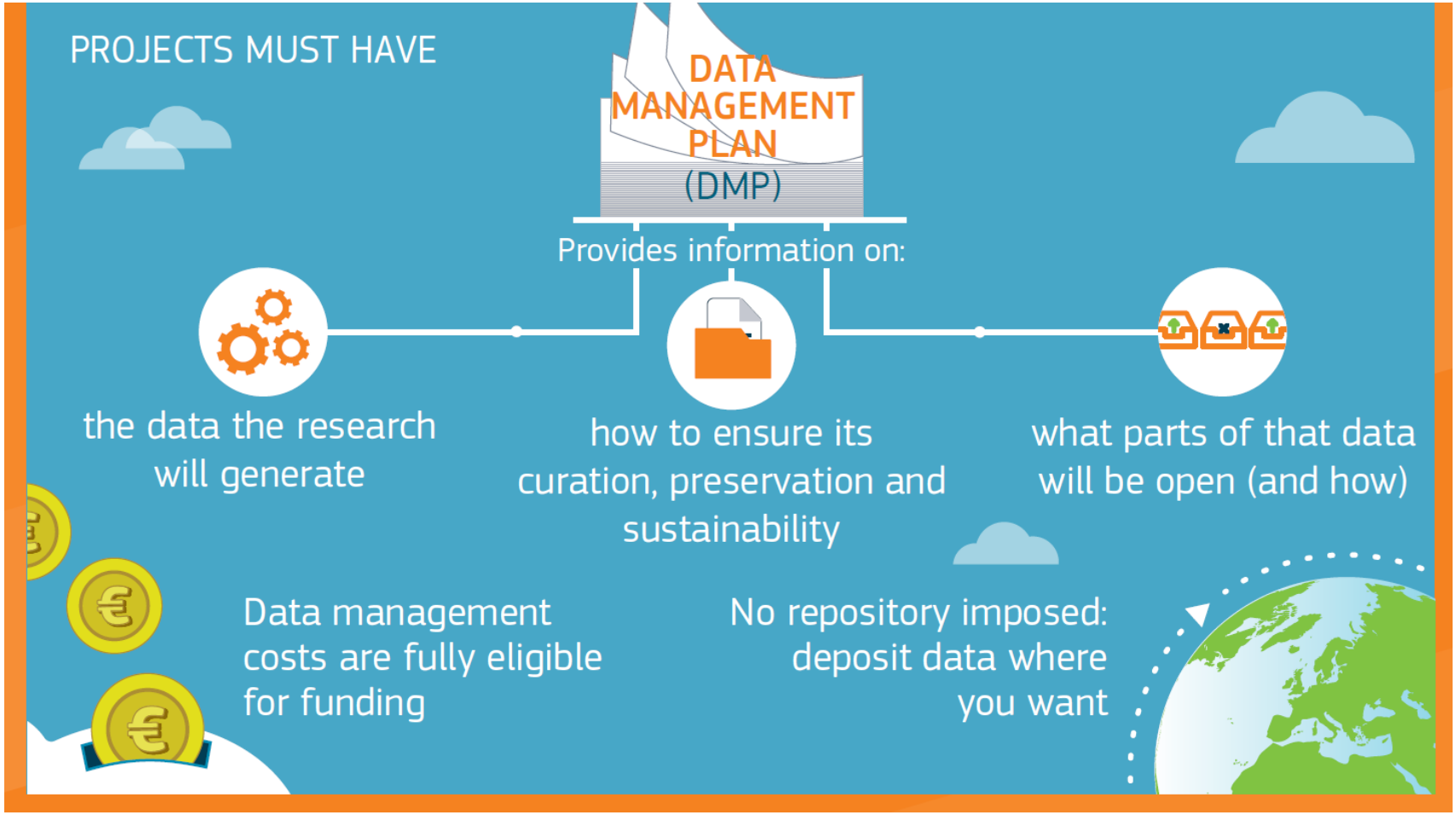
- https://www.force11.org/group/fairgroup/fairprinciples
- Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3:160018 doi: 10.1038/sdata.2016.18 (2016).

EU HORIZON 2020

UNIVERSITÄT BERN

- Open access publications AND
- From 2017: open research data (with possibility to opt out but only with justification)
  - Privacy
  - Intellectual property rights
  - Threatens the main project objectives
- FAIR

EU HORIZON 2020: DMP



SNF on open research data

UNIVERSITÄT BERN

> "The SNSF therefore expects all its funded researchers
  - to store the research data they have worked on and produced during the course of their research work,
  - to share these data with other researchers, unless they are bound by legal, ethical, copyright, confidentiality or other clauses, and
  - to deposit their data and metadata onto existing public repositories in formats that anyone can find, access and reuse without restriction."

SNF: Data Management Plan

UNIVERSITÄT BERN

- Integral part of the proposal ("The proposal can only be submitted once the DMP has been completed.")
- Four sections
  1. data collection and documentation
  2. ethics, legal and security issues
  3. data storage and preservation
  4. data sharing and reuse
- Costs to enable data access eligible
  - if the research data is deposited in recognised scientific, digital data archives (data repositories) that meet the FAIR principles and do not serve any commercial purposes

International Committee of Medical Journal Editors June 6, 2017

Annals of Internal Medicine

EDITORIAL

Data Sharing Statements for Clinical Trials: A Requirement of the International Committee of Medical Journal Editors

The International Committee of Medical Journal Editors (ICMJE) believes there is an ethical obligation to responsibly share data generated by interventional clinical trials because trial participants have put themselves at risk. In January 2016 we published a proposal aimed at helping to create an environment in which the sharing of deidentified individual participant data becomes the norm. In response to our request for feedback we received many comments from individuals and groups (1). Some applauded the proposals while others expressed disappointment they did not more quickly create a commitment to data sharing. Many raised valid concerns regarding the feasibility of the

ples of data sharing statements that would meet these requirements are in the Table.

These initial requirements do not yet mandate data sharing, but investigators should be aware that editors may take into consideration data sharing statements when making editorial decisions. These minimum requirements are intended to move the research enterprise closer to fulfilling our ethical obligation to participants. Some ICMJE member journals already maintain, or may choose to adopt, more stringent requirements for data sharing.

Sharing clinical trial data is one step in the process articulated by the World Health Organization (WHO)

Requirements for publishing a clinical trial in a member journal

- Submission from July 1st, 2018 onwards: Data sharing statement
- Enrollment after January 1st, 2019: Data sharing plan in the trial registration record (clinicaltrials.gov)
- These initial requirements do **not yet** mandate data sharing

Data sharing statement

1. Whether and what individual deidentified participant data (including data dictionaries) will be shared
2. Whether additional, related documents will be available (e.g., study protocol, statistical analysis plan, etc.)
3. When the data will become available and for how long
4. What access criteria data will be shared (including with whom, for what types of analyses, and by what mechanism)

> Examples

Table. Examples of Data Sharing Statements That Fulfill These ICMJE Requirements*

	Example 1	Example 2	Example 3	Example 4
Will individual participant data be available (including data dictionaries)?	Yes	Yes	Yes	No
What data in particular will be shared?	All of the individual participant data collected during the trial, after deidentification.	Individual participant data that underlie the results reported in this article, after deidentification (text, tables, figures, and appendices).	Individual participant data that underlie the results reported in this article, after deidentification (text, tables, figures, and appendices).	Not available
What other documents will be available?	Study Protocol, Statistical Analysis Plan, Informed Consent Form, Clinical Study Report, Analytic Code	Study Protocol, Statistical Analysis Plan, Analytic Code	Study Protocol	Not available
When will data be available (start and end dates)?	Immediately following publication. No end date.	Beginning 3 months and ending 5 years following article publication.	Beginning 9 months and ending 36 months following article publication.	Not applicable
With whom?	Anyone who wishes to access the data.	Researchers who provide a methodologically sound proposal.	Investigators whose proposed use of the data has been approved by an independent review committee ("learned intermediary") identified for this purpose.	Not applicable
For what types of analyses?	Any purpose.	To achieve aims in the approved proposal.	For individual participant data meta-analysis.	Not applicable
By what mechanism will data be made available?	Data are available indefinitely at (*Link to be included*).	Proposals should be directed to xxx@yyy. To gain access, data requestors will need to sign a data access agreement. Data are available for 5 years at a third party website (*Link to be included*).	Proposals may be submitted up to 36 months following article publication. After 36 months the data will be available in our University's data warehouse but without investigator support other than deposited metadata. Information regarding submitting proposals and accessing data may be found at (*Link to be provided*).	Not applicable

* These examples are meant to illustrate a range of, but not all, data sharing options.

How can CTU support you?

- Consulting
- Templates
- Data preparation (Deidentification/anonymization)
- Standards

- Future data sharing platform at the University of Bern (project)