



Paediatric Clinical Research Infrastructure Network PedCRIN

CSA_ H2020-INFRADEV-2016-2017/H2020-INFRADEV-2016-1 (Individual support to ESFRI and other world-class research infrastructures)
Grant Agreement # 731046

Deliverable D4.7 Procedure for patient level data sharing

Date of preparation: 20th July 2017

Working Package: WP4 (ECRIN)

Contact: Jacques Demotes
ECRIN-ERIC Director General
European Clinical Research Infrastructure Network
BioPark, 5-7 rue Watt
75013 Paris, France
Tel: +33 180058646
Jacques.demotes@ecrin.org
www.ecrin.org

Table of Contents

1. Introduction	4
2. Data sharing methods	4
2.1. Public	4
2.2. Quasi-public	4
2.3. Non-public	4
3. Data de-identification methods	4
3.1. Anonymization	5
3.2. Pseudonymisation	5
4. Benefits of data sharing	6
5. Procedure for access to individual patient clinical trial data in PedCRIN	6
5.1. CORBEL Principles of patient data sharing	6
6. Conclusion	7
7. Reference List	7

Abbreviations

ECRIN	European Clinical Research Infrastructure Network
PedCRIN	Paediatric Clinical Research Infrastructure Network
CORBEL	Coordinated Research Infrastructures Building Enduring Life-science Services
IPD	Individual Participant Data

1. Introduction

In recent years, several major organisations have called for increased sharing of the data generated by publicly funded research, including the Organisation for Economic Co-operation and Development (1), the European Commission (2), the National Institutes of Health in the US (3) and the G8 science ministers (4). This trend reflects the growing recognition that: “Publicly funded research data are a public good, produced in the public interest, which should be made openly available with as few restrictions as possible in a timely and responsible manner” (5). With the issue of sharing data from clinical trials in the spotlight, there has been considerable focus on how industry, regulatory bodies, clinical trial funders and sponsors can amend their practices to facilitate clinical trial data sharing. Barriers to data sharing are often presented without discussion of how to overcome them, and although several stakeholders are actively encouraging data sharing activities there are still more opportunities to increase the provision of data to other researchers for further use (8).

2. Data sharing methods

There are three general ways to share data for secondary purposes: public, quasi-public, and non-public. The data needs to be anonymized or pseudonymized in all of the three cases.

2.1. Public

Public data has the least amount of restrictions placed on it. Such data must be anonymized and have no need of reidentification. Public data is available, typically online, for anyone to download either free or for a nominal fee. Many national statistical agencies release census and national survey data as public data (6).

2.2. Quasi-public

Quasi-public data has additional restrictions imposed on it in the form of a “terms of use.” This is a contract that the data recipient signs (or clicks through if it is online). The terms of use often includes a prohibition on attempting to re-identify the data, contacting any of the patients, linking the data with other datasets, and sharing the data with any third party. Also, all data recipients must register so that their identity is known to the data custodian (6).

2.3. Non-public

Non-public data has the most restrictions placed on it. In this case the data recipient would need to sign a full contract that, in addition to the above specifications, includes a prescriptive set of security and privacy controls that the data recipient needs to have in place, such as encrypting their computers and providing privacy training to the analysts who will work with the data. The data custodian may also reserve the right to audit recipients to ensure that they comply with all of the conditions (6).

3. Data de-identification methods

According to Privacy Rule of the Health Insurance Portability and Accountability Act (HIPAA) de-identification of data is defined as protection of all individually identifiable health information that is held or transmitted by a covered entity (7). These are the 18 HIPAA Identifiers that are considered personally identifiable information (name, address (all geographic subdivisions smaller than state, including street address, city county, and zip code), all elements (except years) of dates related to an individual (including birthdate, admission date, discharge date, date of death, and exact age if over 89), telephone numbers, fax number, email address, social security number, medical record

number, health plan beneficiary number, account number, certificate or licence number, any vehicle or other device serial number, web URL, internet protocol (IP) address, finger or voice print photographic image-photographic images are not limited to images of the face, any other characteristic that could uniquely identify the individual. To be considered “de-identified”, all of the 18 HIPAA identifiers must be removed from the data set (7).

Anonymization and pseudonymization are two methods commonly used to de-identify patient data. The aim of de-identification is to obscure the identifiable data items within the persons records sufficiently that the risk of potential re-identification of the subject or a person’s record is minimised to acceptable levels, this will provide effective anonymization (irreversible de-identification) or pseudonymization (reversible de-identification). Although the risk of re-identification cannot be fully removed this can be minimised with the use of multiple pseudonyms. De-identified data should still be used within a secure environment with staff access on a need to know basis. Therefore shared individual participant data from clinical trials used for further scientific research should always be de-identified and either pseudonymised or anonymised

3.1. Anonymization

Anonymisation is a technique applied to personal data to make it, in practice, unidentifiable. It is irreversible, does not require informed consent .It involves the full uncoupling of the data from the original patient. Full (complete, or irreversible) anonymisation involves de-identification and the destruction of any link to an identified or identifiable person via a pseudonym. Effective anonymisation can be applied to a specific dataset, by de-identification and removal of the link to a pseudonym, coupled with the use of new identifiers for individuals. Anonymisation helps organisations to comply with their data protection obligations whilst enabling them to make information available to the public (6;8).

3.2. Pseudonymisation

Pseudonymisation is a procedure by which most identifying fields within a data record are replaced by one or more artificial identifiers. This procedure is reversible and require informed consent for data reuse. There can be a single pseudonym for a collection of replaced fields or a pseudonym per replaced field. The purpose of pseudonymisation is to render the data record less identifying and, therefore, to lower customer or patient objections to its use. Data in this form are suitable for extensive analysis and processing but could allow a trace back to the original identity of the subject involved (6;8).

The choice of which data fields are to be pseudonymised is partly subjective, but should include all fields that are highly selective, such as, for example, personal health identification number. Less selective fields, such as birth date or postal code are often also included because they are usually available from other sources and therefore make a record easier to identify by cross-referencing. Pseudonymizing these less identifying fields removes most of their analytic value and should therefore be accompanied by the introduction of newly derived and less identifying forms, such as age or a larger postal code region.

The advantage of sharing pseudonymised data is that, if the secondary user discovers good reasons for clarifying, expanding or matching some of the data, or even for further investigations with some of the source population, they can contact the holders of the pseudonymous data and discuss if and how this might be achieved, because the individual participants are still (indirectly) identifiable. This does not mean that identifiable or identifying information would be transferred to a secondary user, unless there was explicit consent from the participant for this to happen (though this seems unlikely to be given). It only means that if a case can be made for identifying the

individuals in the data set it is at least possible to discuss the possibilities of doing this, including possibly returning to the individuals concerned to request additional consent or informed consent when original data were collected for research purposes.

4. Benefits of data sharing

Sharing Individual Participant Data (IPD) from clinical trials offers numerous well recognized advantages that can advance clinical research and benefit patients (6). The expected benefits from sharing individual patient data for health research purposes include:

1. It ensures accountability in results and that reported study results are valid,
2. It allows researchers to build on the work of others more efficiently and to perform individual patient data meta-analyses to summarise evidence
3. it decreases the burden on research subjects through the reuse of existing data.

5. Procedure for access to individual patient clinical trial data in PedCRIN

Providing the scientific community with access, upon request, to access to individual patient-level clinical trial data is part of the European Clinical Research Infrastructure Network (ECRIN) eligibility criteria, assessed during the evaluation of projects asking for support to trial management. This eligibility criterion has been used for the selection of projects for Paediatric Clinical Research Infrastructure Network (PedCRIN) funding. A single stage call for funding was launched in Feb 2017 to provide operational support to selected pilot paediatric clinical trials and to test the developed tools. In addition to the tool development one objective is to develop a framework in which, ultimately, all of the patient level data from these pilot paediatric clinical trials becomes available to those who can demonstrate they can make appropriate use of it. However, there is currently no established procedure (who should request data, what should be the content of the protocol for re-analysis / meta-analysis, who should be the data custodian providing access etc.).

To overcome this barrier PedCRIN members (WP3) are establishing a procedure for patient level data in paediatric and neonatal trials based on the principles and practical recommendations set by the Horizon2020-funded project CORBEL (Coordinated Research Infrastructures Building Enduring Life-science Services) WP3 coordinated by the ECRIN. The CORBEL principles and practical recommendations for providing access to individual participant data from clinical trials focused particularly on non-commercial European trials (8).

5.1. CORBEL Principles of patient data sharing

Ten principles emerged from the CORBEL consensus process (Table 1), representing what the task force saw as the fundamental requirements for any framework for the sharing and re-use of clinical trials data(8)

Table 1: Principles of Data Sharing in Clinical Trials(8)

No.	Principles of data sharing
1.	The provision of individual-participant data should be promoted, incentivised and resourced so that it becomes the norm in clinical research. Plans for data sharing should be described prospectively, and be part of study development from the earliest stages.
2.	Individual-participant data sharing should be based on explicit broad consent by trial participants (or if applicable by their legal representatives) to the sharing and re-use of their data for scientific purposes.

3.	Individual-participant data made available for sharing should be prepared for that purpose, with de-identification of datasets to minimise the risk of re-identification. The de-identification steps that are applied should be recorded.
4.	To promote inter-operability and retain meaning within interpretation and analysis, shared data should, as far as possible, be structured, described and formatted using widely recognised data and metadata standards
5.	Access to individual-participant data and trial documents should be as open as possible and as closed as necessary, to protect participant privacy and reduce the risk of data misuse.
6.	In the context of managed access, any citizen or group that has both a reasonable scientific question and the expertise to answer that question should be able to request access to individual-participant data and trial documents.
7.	The processing of data access requests should be explicit, reproducible, and transparent but, as far as possible, should minimise the additional bureaucratic burden on all concerned.
8.	Besides the individual-participant datasets, other clinical trial data objects should be made available for sharing (e.g. protocols, clinical study reports, statistical analysis plans, blank consent forms), to allow a full understanding of any dataset.
9.	Data and trial documents made available for sharing should be transferred to a suitable data repository, to help ensure that the data objects are properly prepared, are available in the longer term, are stored securely and are subject to rigorous governance.
10.	Any dataset or document made available for sharing should be associated with concise, publicly available and consistently structured discovery metadata, describing not just the data object itself but also how it can be accessed. This is to maximise its discoverability by both humans and machines.

6. Conclusion

The patient level data sharing recommendations set by the Horizon2020-funded project CORBEL would not only assist PedCRIN members towards establishing a procedure for patient level data in paediatric and neonatal pilot trials (PedCRIN funded trials) but will also be a window of opportunity for implementing and testing in practice.

7. Reference List

- (1) The Organisation for Economic Co-operation and Development. OECD Principles and guidelines for access to research data from public funding <http://www.oecd.org/science/sci-tech/38500813.pdf> 2007.
- (2) C(2012) 4890 final Commission recommendation of 17/7/2012 on access to and preservation of scientific information. European Commission. Available at http://ec.europa.eu/research/science-society/document_library/pdf_06/recommendation-access-and-preservation-scientific-information_en.pdf 2012.
- (3) National Institutes of Health Plan for Increasing Access to Scientific Publications and Digital Scientific Data from NIH Funded Scientific Research. NIH. Available at <https://grants.nih.gov/grants/NIH-Public-Access-Plan.pdf> 2015.
- (4) G8 Science Ministers Statement. Available at <https://www.gov.uk/government/news/g8-science-ministers-statement> 2013.
- (5) RCUK Common Principles on Data Policy. Research Councils UK. [Available at http://www.rcuk.ac.uk/research/datapolicy/](http://www.rcuk.ac.uk/research/datapolicy/) 2015.
- (6) El Emam K, Rodgers S, Malin B. Anonymising and sharing individual patient data. *BMJ* 2015 Mar 20;350:h1139.

- (7) Health Insurance Portability and Accountability Act, P. L. 104–191, 110 Stat 2023 (1996).
- (8) Christian Ohmann, Rita Banzi, Steve Canham, Serena Battaglia, Mihaela Matei⁴, Chris Ariyo, Lauren Becnel, Barbara Bierer, Sarion Bowers, Luca Clivio², Monica Dias, Christiane Druml, H el ene Faure, Martin Fenner, Jose Galvez, Davina Gherzi, Christian Gluud, Trish Groves, Paul Houston⁶, Ghassan Karam, Dipak Kalra, Rachel Knowles, Karmela Krleza-Jeric, Christine Kubiak⁴, Wolfgang Kuchinke, Rebecca Kush, Ari Lukkarinen⁵, Pedro Marques, Andrew Newbigging, Jennifer O'Callaghan, Philippe Ravaud, Irene Schl under, Daniel Shanahan, Helmut Sitter, Dylan Spalding, Catrin Tudur Smith, Peter Van Reusel⁶, Evert-Ben Van Veen, Gerben Rienk Visser, Julia Wilson, Jacques Demotes-Mainard⁴: Sharing and re-use of individual participant data from clinical trials: principles and recommendations. CORBEL (Coordinated Research Infrastructures Building Enduring Life-science Services). *Manuscript submitted for publication, 2017.*