RESTART Data Sharing

Overview

An anonymised version of the dataset used for analysis with individual participant data and a data dictionary will be available for sharing using a controlled-access model. Researchers can apply for use from 22nd May 2020, via https://datashare.is.ed.ac.uk/handle/10283/3265. Written proposals will be assessed by members of the RESTART trial steering committee and a decision made about the appropriateness of the use of data. A data sharing agreement will be put in place before any data are shared.

Process

It was agreed with the chief investigator, Rustam Al-Shahi Salman, the data to be shared would consist of the variables to recreate the tables in the main results Lancet publication (https://www.sciencedirect.com/science/article/pii/S0140673619308402?via%3Dihub) and the imaging sub-group analyses of the trial, published in the Lancet Neurology (https://www.sciencedirect.com/science/article/pii/S147444221930184X?via%3Dihub).

Anonymising the study dataset

A reduced dataset containing only the variables required to repeat the main analyses was created.

This resulted in the removal of: all direct identifiers such as name, initials etc. and superfluous data, such as audit trail data.

All dates relating to individuals were classed as direct identifiers. For anonymisation purposes, date of randomisation was used as a reference date for each participant, classed as day 0. Complete dates (i.e. those where a day, month and year are provided) were modified to be relative to day 0.

The unique identifying number (e.g. subject number) was recoded using random number generator methodology which ensured reproducibility and linkage to the original unique identifier. The link has been maintained in case of any queries relating to the anonymised dataset from secondary researchers.

Indirect identifiers, sex and ethnicity, remain in the anonymised dataset. Ethnicity was grouped into white versus other to reduce the risk of identification.

Some variables in the dataset may pose a minimal risk of identification when used in combination with others as they can result in identifying a small number (<3) of distinctive participants based on their baseline characteristics, treatment allocation and outcomes.

Assessing the possibility of identification

We considered what means are reasonably likely to be used to identify the individual taking into account all objective factors, such as:

- the costs and amount of time required for identification;
- the available technology at the time of the processing; and
- likely technological developments,

and believe the risk of identification is minimal. Precautions to minimise the risk of identification have been taken which include sharing a reduced dataset containing only the variables required to repeat the main analyses. Another precaution is using a restricted access model, where the requester must prove that they are a bona fide researcher with a sensible question to answer and appropriate data-sharing agreements are to be signed. Further reduction of the dataset would not enable someone to repeat the final analyses.

Study data pack

In addition to the anonymised dataset, the following documents are to be shared as part of the study data pack:

1. Study protocol, including all previously released versions
2. Case Report Forms. Annotated with the data variable names for clear identification of the dataset
3. Data dictionary
4. Statistical Analysis Plan
5. Primary publication

Releasing the data

It has been confirmed the anonymisation process has not impacted on the usefulness of the data and the primary analysis can be repeated using the data to be shared.

The study pack will be uploaded to Edinburgh DataShare, the University of Edinburgh's digital respository of research data (https://datashare.is.ed.ac.uk/handle/10283/3265). The study pack will be held under embargo, requiring secondary users of the data to apply to access the data. This controlled access model further reduces the risk of reidentification.