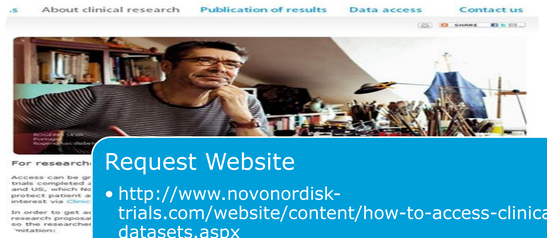
De-identification of clinical trial data at Novo Nordisk A/S to enable secondary use of anonymized* data for research

In order to facilitate the use of data from our clinical research in further scientific investigations, Novo Nordisk has established a new system for sharing of data with external researchers in response to legitimate scientific requests. Via this system external researchers can submit their requests for access to anonymised* datasets from Novo Nordisk sponsored clinical trials.

In case of approval, Novo Nordisk provides t de-identified trial data to protect patient anonymity together with the redacted Clinical Study Report, thus enabling the researcher to complete the proposed analysis.

Secondary use of clinical data for research is not limited to data access requests received via our Data Access Website, but also covers sharing de-identified clinical trials to be used in clinical databases for specific therapeutic areas (e.g. Haemophilia trials, Hypoglycaemia trials,…) where Novo Nordisk A/S and other pharmaceutical companies contribute to the database by sharing de-identified clinical trials. Today Novo Nordisk A/S is participating in 2 databases for research, the first one is TransCelerate Placebo/Standard of Care, and the second one is a newly started project to share Hypoglycaemia trials together with other big pharma companies.

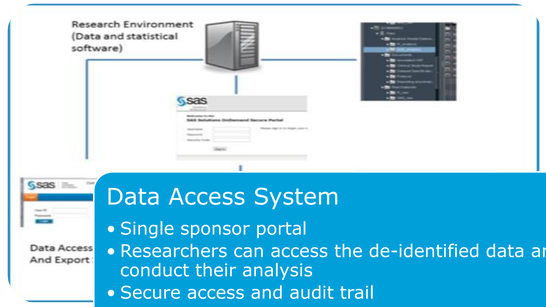


Request Website

- http://www.novonordisk-trials.com/website/content/how-to-access-clinical-trial-datasets.aspx
- Researchers can submit their data access request from this website

Independent Review Board

- Data Access Request
- Data Sharing Agreement



Data Access System

- Single sponsor portal
- Researchers can access the de-identified data and conduct their analysis
- Secure access and audit trail

Approach

Expert Determination method, data sharing agreement and risk measurement.

Metadata controlled de-identification and documentation.

Both horizontal and vertical (column and row) de-identification.

Only requested trials are de-identified.

A set of de-identification rules defined.

De-identification rules similar to other companies rules and compliant with guidance from TransCelerate and other organisations.

In-house developed SAS solution.

In-house developed risk-assessment tool.

Generic code that can de-identify trials from 5 different data models.

Design

Simplicity - one macro can de-identify any trial from 5 different data models.

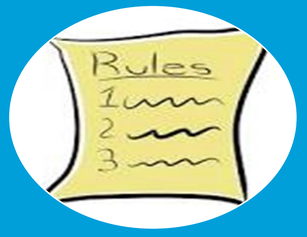
Triple Nested DO-loops to implement the de-identification rules :

First loop: Libraries.

Second loop: Tables.

Third loop: Columns.

The design enables de-identification of Derived data, Study data and other data folders.



Rules

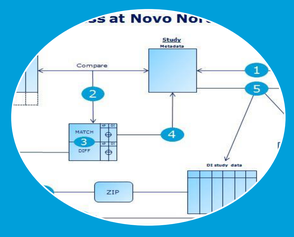
14 rules for vertical de-identification (columns)

Randomization, offsetting, grouping, suppression, computation, redaction, …

3 rules for horizontal de-identification (rows)

Masking, suppression, …

Additional rules can be added if needed



Process

A metadata controlled process to control both de-identification and documentation.

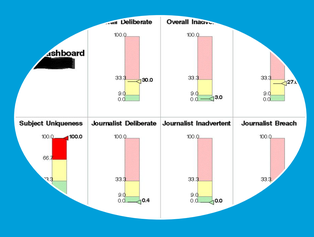
Master metadata for each data model is used to get de-identification rules automatically. Master metadata can also be updated by trial specific metadata.

Consistency check report to secure consistency between trial data and de-identified data.

Documentation

The de-identification process is documented by creating define.xml files that document the used de-identification rules and metadata for the datasets.

SAS Clinical Standard Toolkit is used for creating the define.xml files.



Risk Assessment

Our expert determination approach is based on risk assessment to ensure that we meet the specified risk threshold and to document the considerations used in the de-identification process.

The risk assessment tool creates a report and a risk dashboard that gives a good overview of the calculated risks.

* Anonymised subject data is irreversible de-identified data (where the key, randomization seed and the link between original data and de-identified data are deleted to prevent traceability to subjects) of already anonymous subject data.

Poster by Adel Salem