

# **Vivli and Microsoft Innovation Challenge**

Adel Salem  
Novo Nordisk A/S  
21 June 2019

# Challenge

- Anonymizing rare disease trial while balancing privacy and data utility.

# Approach

- Anonymization of rare disease trial is done case by case after considering the following considerations:
- population size
- quasi identifiers needed for the data requester to do the analysis
- Release context (public, trusted recipient, mitigating controls, data sharing agreement, ...etc)
- The possibility to anonymize the data to an accepted risk of re-identification.

# The tool

- A self developed SAS macro with generic code that can anonymize any data regardless of data model or data structure.
- A self developed tool to measure the residual risk after anonymization based on three different attack scenarios:

# De-identifications

- Direct Identifiers recoded
- Region information is blanked: assumed not necessary for the analysis.
- Dates offset
- Free text columns blanked to avoid revealing Personal information, if the text is not required/necessary.
- Trial ID is recoded to minimize risk of re-identification
- Making sure of enough subject with the same equivalence class.
- Measure residual risk

# Risk assessment

