



**Dementias**  
Platform<sup>UK</sup>  
Medical Research Council

# Introducing the DPUK synthetic dataset

# What is the synthetic dataset?

*The synthetic dataset is invented data in the format of a typical cohort dataset on the DPUK Data Portal.*

- It does not contain real data.
- The data are invented and do not intended to represent any particular study or group of cohort participants.
- As the data are randomised, they are not owned by anyone.

# Why has it been created?

- A researcher does not need any permission to be able to use these data.
- The dataset can be accessed by anyone.
- The dataset is immediately available.
- The data contained in the dataset are of the same types as a researcher can expect to see in a 'real' cohort dataset.

# Why is it useful?

- The data enables researchers to explore the Data Portal's capabilities before they apply for data for their studies.
- Researchers can analyse the data without having to obtain permission first.
- Researchers can review how to work on the Data Portal before they apply for a study and before they receive 'real' data.
- Researchers can develop code to use in their project and try other methods, such as machine learning.

# How can I access it?

- The dataset can be accessed on the DPUK Data Portal.
- It is available in a separate sub-directory, that all portal users have access to.
- It is in S:\.....\..... .
- Researchers can use the software on the Data Portal to analyse the dataset.
- Please don't publish your results!