We have created an illustrative NCHDA dataset of 500 patients for users to test and explore the code.The illustrative dataset is fully anonymised and not intended to represent real patients.

To ensure full anonymisation and that the data does not correspond to any real patient, we followed these steps:

1. Randomly selected 500 patients from the raw data we received from NICOR, retaining only the first chronological record for each patient.
2. Generated a synthetic version of each record with the following fields. For example, the record for patient A is shown below, with B, C, D… representing other individual patients.

* **Pseudo study ID**: just a numeric identifier from 1 to 500, for both record and patient identifier (as in this example we just have one record per patient)
  + LAUNCHESpatID: patient identifier
  + LAUNCHESrecID: record identifier
* **Date of Birth**:
  + dob.year: actual year of birth from patient A
  + dob.month: randomly assigned month from January to December
* **Age at Operation and Discharge**:
  + Raop: actual age at operation from patient A with a random age offset
  + Raod: age at operation plus a hospital stay duration, randomly sampled from the empirical distribution of lengths of stay in the entire dataset.
* **Discharge Status**

X4.03.Discharge.Status: randomly assigned with a 1.8% probability of “D. Died in hospital” 0.2% probability of “123333. Death within 30 days of procedure”, and 97% probability of “A. Alive”.

* + Paod (age at death): assign as age at discharge “Raod” if died in hospital or within 30 days of procedure.
* **ONS linkage – life status**
  + Life\_status: assigned as "death" if indicated in Discharge Status. If discharged alive, a life status is randomly assigned with a 2% probability of death and 98% probability of being alive.
  + Age\_life\_status: if alive, assign the age at 2022 using year of birth. If deceased, for those who died at discharge, age at discharge is used; for those who died after discharge, a random age at death is assigned by adding a randomly generated time to the age at discharge.
* **Admission/Demographic characteristics**
  + X1.01.hospital: from another patient (e.g., patient B)
  + X1.07.Gender: from another patient (e.g., patient C)
  + X1.08.Ethnic: from another patient (e.g., patient D)
  + X1.09.Patient.Status (eg NHS or private): from another patient (e.g., patient E)
  + country: from another patient (e.g., patient F)
  + IMD2004-2019 decile: from another patient (e.g., patient G)
* **Diagnosis fields**
  + X2.01.Diagnosis, Diagnosis.2 to Diagnosis.25..to.32: from another patient (e.g., patient H)
* **Procedure fields**
  + X3.07.ProcedureType (type of procedure) and X3.09.Operation.Performed, Procedure.2 to Procedure.7 (detailed procedure code): from the patient we used to generate diagnosis fields
  + X2.02.Previous.Procedure: empty because we only retain the first record
* **Comorbidity fields**
  + X2.06b.Comorbidity.present: Randomly assigned, with a 22% probability of comorbidity being present and a 78% probability of no comorbidity.
  + X2.07.Comorbid.Conditions: if comorbidity present, codes are randomly selected from the comorbidity code list.
* **Other fields**
  + X2.04.AntenatalDx (any evidence of antenatal detection): from another patient (e.g., patient J)

Please note that not all variables from the original NCHDA dataset are included. For detailed descriptions of each variable and the complete list, refer to the NCHDA Data Manual: https://www.nicor.org.uk/~documents/data-sets/supporting-data-set-documentation/congenital-heart-disease-1/nchda-manual-v6-1-march-23-revision-23-24-final/?layout=file.