# READ ME

There are 9 sentinel Congenital Heart Disease (CHD) diagnoses and one should run the programming for each CHD diagnosis in the following order (decreasing order of complexity):

* Hypoplastic left heart syndrome (HLHS)
* Functionally univentricular heart (FUH)
* Transposition of the great arteries (TGA)
* Pulmonary atresia (PA)
* Atrioventricular septal defect (AVSD)
* Tetralogy of Fallot (TOF)
* Aortic stenosis (AOS)
* Coarctation (COA)
* Ventricular septal defect (VSD)

For each CHD, the main R file is “CHDtype\_main file.R”, which calls functions from generic\_modules.R and CHDtype \_modules.R.

**Before running the code, one needs to carefully read the rules for individual CHD, which include all code lists for inclusion and exclusion, assigning pathways & diagnosis subgroup, and handling data missing/errors, etc.**

**For enquiries, qi\_huang@ucl.ac.uk**

## Data sources

The core dataset is the National Congenital Heart Disease Audit (NCHDA) with ONS death registration data linked. NCHDA is run by the National Institute for Cardiovascular Outcomes Research (NICOR), which collects data to assess patient outcomes after therapeutic paediatric and congenital cardiovascular procedures (surgery, transcatheter, and electrophysiological interventions). CHD and paediatric cardiac operations are very diverse and complex, and are described in NCHDA using a special coding scheme from the European Paediatric Cardiac Code (EPCC), itself a derived Short List of the International Paediatric and Congenital Cardiac Code (www.ipccc.net). One can download the NCHDA data manual and EPCC codes at NICOR’s website (https://www.nicor.org.uk/datasets/supporting-data-set-documentation). To note, as the NCHDA is a procedure-based dataset, patients who did not undergo any surgical or interventional cardiac procedures do not appear in the dataset.

### Important variables

We use the following important variables to assign the sentinel CHD diagnosis, subgroups, treatment pathway and suspected missing data.

* patid: patient-level identifier
* ageatop/ageatdis (record-level, derived): age at operation/discharge, which was computed by the difference between the date of operation/discharge and the birth date.
* patentry (record-level, derived): one patient may have multiple procedure records. We order the entry of each record by age at operation, and if multiple records have the same age at operation, we use age at discharge to order the entry.
* diagcode1-29 (record-level, derived): EPCC codes in diagnosis filed (cleaned, contain the 6 digits EPCC Code only)
* comorbidity1-16 (record-level, derived): codes in comorbidity filed (cleaned, contain the 6 digits EPCC Code only)
* proccode1-7 (record-level, derived): codes in procedure filed (cleaned, contain the 6 digits EPCC Code only)
* prevproccode1-26 (record-level, derived): codes in previous procedure filed (cleaned, contain the 6 digits EPCC Code only)
* sp\_allocation (record-level, derived): specific procedure algorithm allocation of procedure type (version 8.05, used in NICRO. See R code in https://github.com/fespuny/CORUcode/tree/main/AA%20SP%20-%20R%20Code)
* aa\_allocation (record-level, derived): activity algorithm allocation of procedure type. The algorithm is developed by NICOR (version 8.03, used in NICRO. See R code in https://github.com/fespuny/CORUcode/tree/main/AA%20SP%20-%20R%20Code)
* InterType (record-level, derived): intervention type derived from aa\_allocation

The labels are:

1: Surgery and hybrid (aa\_allocation 1:bypass ,2:non-bypass,3:hybrid)

2: Interventional catheters and ep (aa\_allocation 6: icd:non-surgical, 7: pacemaker:non-surgical, 8: ep:non-surgical, 9: intervention:non-surgical)

3: Excluded (aa\_allocation 4: vad, 5: unallocated-ecmo, 10: diagnostic:non-surgical, 11: unallocated, 12: primary ecmo)

* lastknownstatus/ageatlastknowstatus (patient-level, derived. see in data\_processing\_for\_NCHDA.R): vital status (0 censored/1 death) and age at vital status, which were computed from two data sources: NCHDA and ONS death registration.

Patient vital status (dead or alive) was provided at the point of hospital discharge by NCHDA, who obtained this information from treating centres. The age at death for any patient who had died was taken from death certification data provided by the ONS. For surviving patients, we received from ONS their age when this status was confirmed. Any patients that were discharged alive and who had missing life status with ONS were deemed lost to follow up and were censored at their most recent discharge age provided by NCHDA.

* lastknownstatus\_NCHDA/ageatlastknowstatus\_NCHDA (patient-level, derived in data\_processing\_for\_NCHDA.R): vital status (0 censored/1 death) and vital at survival status as of the last date of NCHDA dataset.

There may be a time gap between the ONS confirmation date and the last date of the NCHDA dataset (due to the time required for regulatory processes), so procedures undertaken during this time gap may have been missed. We therefore used patients’ vital status as of the last date of NCHDA to calculate the follow-up time and status of procedures.

## Code example: HLHS

For HLHS, running the code in HLHS\_main\_file.R will result in the following.

### Step 1 HLHS inclusion and exclusion module: identify patients with evidence of HLHS

#### Input dataset

NCHDAdata (Initial NCHDA cohort in record-level)

#### Derived variables

HLHSpat (patient-level): marker of HLHS patients, 0: No /1: Yes

#### Output dataset

HLHSdata\_raw (HLHS cohort in record-level, raw data)

### Step 2 HLHS pathway module: assign each patient's procedure records to a category and diagnostic subgroup

#### Input dataset

HLHSdata\_raw from step 1

#### Derived variables

* CatProc (record-level): Category of each procedure

A generic label for CatProc was used for all CHDs:

*0: pre-pathway*

*1: palliative stage 1 (pathway)*

*2: single ventricle palliative stage 2 (pathway)*

*3: single ventricle palliative stage 3 (pathway)*

*4: reparative procedure (pathway, empty in HLHS)*

*5: other pathway (Hybrid procedure in HLHS)*

*6: heart transplantation*

*7: reintervention/off-pathway*

*8: excluded (non-contributory record)*

Of note, procedure records with CatProc 1,2,3,4, 5 and 6 were considered as pathway.

* patAge\* (patient-level): age at pathway. For example, patAge1 and patAge2 are age at palliate stage 1 and 2, respectively.
* ProcSeq (patient-level): procedure sequence (pathway and heart transplantation). For example “1236” means patient had palliative stage 1, 2, 3 and heart transplantation.
* diagsubgroup (patient-level): diagnosis subgroup. HLHS has only 1 group as HLHS.

#### Output dataset

HLHSdata\_raw (derived variables added, record-level data)

### Step 3 HLHS further processing module: identify patients who met the specific violation rule and who met the flag to center rule.

#### Input dataset

HLHSdata\_raw from step 2

#### Sub module:

* Generic violation module
* HLHS specific violation module
* Generic suspected missing/miscoded data module
* HLHS specific suspected missing/miscoded data module
* Generic minor data error data module
* HLHS specific minor data error data module

#### Derived variables

* pat\_violation/ pat\_violationinf (patient-level): marker of patients who met the violation rule (0: No /1: Yes) and the reason. They will be removed from the study cohort.
* Flagcenter\_missingdata/ Flagcenter\_missingdatainf (patient-level): marker of patients suspicious missing or unusual procedure sequences (0: No /1: Yes) and the reason. They will remain in the study cohort. In future routine monitoring all such patients will be flagged with the treating centres for correction.
* Flagcenter\_minor/ Flagcenter\_ minorinf (patient-level): marker of patients with minor data errors (0: No /1: Yes) and the reason. They will remain in the study cohort. In future routine monitoring all such patients will be flagged with the treating centres for correction.

#### Output dataset

HLHSdata\_final (derived variables added, and patients who met the violation rule has been removed. record-level data)

### Step 4 Descriptive analysis module (generic): assign pathway management route, comorbidity and compute number of interventions, pathways and reinterventions for each patient.

#### Input dataset

HLHSdata\_final from step 3

#### Derived variables

* patpremat (patient-level): marker of pre-term birth (0: No /1: Yes)
* patCongComorb (patient-level): marker of congenital comorbidity (0: No /1: Yes)
* patDowns (patient-level): marker of Down syndrome (0: No /1: Yes)
* pathas\*(patient-level): marker of patients who had certain pathway procedure in records (0: No /1: Yes)
  + pathasStage1: patients had palliative stage 1
  + pathasSVstage2: patients had palliative stage 2
  + pathasSVstage3: patients had palliative stage 3
  + pathasBVpathway: patients had reparative procedure (none in HLHS)
* num\* (patient-level): number of interventions for each patient
  + numPrePath: number of pre-pathway
  + numallReint: number of reintervention (any type)
  + numHSReint: number of reintervention (surgery or hybrid)
  + numCEreint: number of reintervention (catheter or electrophysiology )
  + numIntby1: number of all cardiac interventions (all procedures except for those marked as excluded in var. CatProc) by age 1-year
  + numPathwayby1: number of pathway (procedure marked as 1,2,3, 4, and 5 in var CatProc) by age 1-year
* Status\_\*(patient-level): reintervention status
  + Status\_allReint: status of reintervention (any type)

Labels:

0: censored without reintervention

1: occurrence of reintervention

2: Death or heart transplantation without reintervention

* + Status\_HSReint: status of reintervention (surgery or hybrid, labels see above)
  + Status\_HSReint: status of reintervention (catheter or electrophysiology, labels see above)
* pathwaytype (patient-level): whether the patient was managed under a single ventricle pathway or a biventricular reparative pathway. Of note, HLHS is exclusively single ventricle.

#### Output dataset

HLHSdata\_final (derived variables added, record-level data)

### Step 5 Save patient level data module (generic): create a patient-level dataset that includes all the patient-level variables derived from previous steps.

#### Input dataset

HLHSdata\_final from step 4

#### Output dataset

HLHSpatientleveldata (patient-level data)

### Step 6 compute metrics module (generic): compute survival and reintervention metrics.

#### Input dataset

HLHSpatientleveldata from step 5

#### Output aggregated results

* survival metrics: survival rates at 1-year, 5-year and 10-years (using Kaplan-Meier estimator)
* reintervention metrics: cumulative incidence rates of reintervention rates at 1-year, 5-year and 10-years (using competing risk analysis, accounting for death without reintervention as competing events).