UDMI IPDMA Codebook

This document provides an overview of the analysis code we have produced for the collaborative IPDMA. The code contains a script to modify your dataset variables into the correct format and will run self-contained scripts which will place the output from regression models directly into folders [Tables / Figures].

These folders can be directly shared with Andrew in Edinburgh (via secure file transfer or otherwise – a.r.chapman@ed.ac.uk) and **contain no individual patient level data**.

Steps to follow:

Please create a new .Rproj in R Studio before running any code (see below)

https://r4ds.hadley.nz/workflow-scripts.html#rstudio-projects

File -> New Project -> New Directory -> New Project

Name the project after your cohort (e.g. BACC, APACE, Karolinska). Once you have created your R.proj, create a folder inside the directory called Code and save all .R code files in there.

Open the "prepare-data-ipdma.R" file.

Please review and modify according to your cohort based on the instructions (e.g. change dataset name, rename variables, create new variables, specify whether you have heart failure follow-up data etc).

It is key that the following variables are renamed exactly as coded and you use the name in bold:
- **age** (continuous variable)
- **sex** (binary variable)
- **PrevIHD** (previous ischemic heart disease, binary variable)
- **DM** (diabetes mellitus, binary variable)
- **gfr** (glomerular filtration rate, CKD-EPI formula, continuous variable)
- **ischemic_ecg** (defined as ST-depression or T wave inversion, binary variable)
- **T1MI** (type 1 myocardial infarction, binary variable)
- **T2MI** (type 2 myocardial infarction, binary variable)

These variables are later used in the regression models, except T1MI and T2MI which are used for creating the udmi4 variable.

You will find an "only_injury" object. The default is FALSE. If your cohort only includes patients with myocardial injury (i.e. OLMSTED County [Gulati et al], NY HARBOR [Smilowitz et al] or Long Term Outcomes [Chapman et al] change to TRUE. The levels of the udmi4 variable will change accordingly for the regression models.

You will find an "HF_available" object. The default is TRUE. If you have heart failure data, leave it as TRUE and rename/mutate HF variables. If you do not have heart failure data, then change to FALSE and ignore.

At the end of the file, if you have chosen "only_injury" <- FALSE, your data frame will be subset to remove the non-myocardial injury group for a sensitivity analysis.

Save this file.

Go to "flowchart-ipdma.R" file. Please insert start date of recruitment for "time1" and end of recruitment for "time2" and adapt the text in "txt1", "txt1_side" (if you had exclusions) and "txt2".

Save this file.

Go to "table1-ipdma.R" file and adapt with your variable names. The following variables are included as default, but if you do not have all, please insert the variables you have. You will need to adapt the row lines when adding the header rows and colors.

"age"
"sex"
"early_presenter" -> early presenter defined as time since chest pain onset ≤3h
"Time_CPO" -> time since chest pain onset
"HTA" -> hypertension
"DL" -> dyslipidemia
 "DM" -> diabetes mellitus
"Smoking" -> active smoking
"PrevIHD" -> previous ischemic heart disease
"Prev_AMI" -> previous myocardial infarction
"PrevPCI_CABG" -> previous revascularization (PCI or bypass)
"Prev_stroke" -> previous stroke
"stdep" -> ST depression in the ECG
"twave" -> T wave inversion in the ECG
"LBBB" -> left bundle branch block
"RBBB" -> right bundle branch block
"gfr" -> glomerular filtration rate
"antiplatelet_therapy" -> antiplatelet therapy on admission
" anticoagulant_therapy " -> anticoagulant therapy on admission
"betablocker_therapy" -> beta blocker therapy on admission
"ACEIs_ARBs" -> ACE inhibitor or angiotensin receptor blocker on admission
"statin_therapy" -> statin on admission
"calciumchannel_therapy" -> Calcium antagonist on admission
"gtn_therapy" -> GTN on admission

Save this file.

You do not need to adapt any other source file.

Please go to **"main-ipdma.R" file**

For Olmsted or NY-Harbor please go to **"main-ipdma-only-injury-cohorts.R"**

The first function (check_packages) will check whether you have all required packages. If you do not have one, this will be reported in the console and you will need to install.

Once you have all required packages installed, source the "library-ipdma.R" file.

Insert the name of your cohort in the object cohort (without spaces- e.g replace APACE)
cohort <- glue("APACE-{date}")

Source:
source("Code/prepare-data-ipdma.R") #code for preparing data
source("Code/flowchart-ipdma.R") #Flowchart following consort recommendations
source("Code/table1-ipdma.R") #Table 1
source("Code/crude-events-table.R") #obtain table with crude events at 30d and 1y
source("Code/functions-all-ipdma.R") #functions needed for formal regression analysis. Source only once

The next line of code is for choosing the follow-up time (1y or 5y). If you only have 1y follow-up data, leave it as it is (default 1).

If you have 5 years follow-up data, **please run everything two times**: first with 1y [1] and then with 5y [2]

```
fu_time <- c("1y", "5y")[1] #define FU time (select 1y with [1] or 5y with [5])
if (fu_time == "1y") {
  time <- "days.to.mi.death.1y" #introduce your 1y time variable (time to AMI or CV death)
  status <- "status.1y" #introduce your 1y status variable (three level variable: 0 censored, 1 MI or CV death, 2 non CV death [competing event])
} else if (fu_time == "5y") {
  time <- "days.to.mi.death.5y" #introduce your 5y time variable (time to AMI or CV death)
  status <- "status.5y" #introduce your 5y status variable (three level variable: 0 censored, 1 MI or CV death, 2 non CV death [competing event])
}
```

With this you will define your event and time variable. It is paramount that you have correctly renamed/created these variables in the "prepare-data-ipdma.R" file.

source("Code/csHR.R") #code for running cause-specific Cox model

If your cohort includes patients *without* myocardial injury, then the object "only_injury" will be FALSE and the chunk of code below will run the main analysis (cause-specific Cox model) in the subsetted data frame (df_sa) as a sensitivity analysis.

If your cohort only includes patients ***with*** myocardial injury (i.e OLMSTED, NY HARBOR< LONG TERM), then your object "only_injury" will be TRUE and this code will not run.

```r
if (only_injury == FALSE) {
  run_cox_models(df = df_sa, time = time, status = status, status_code = 1,
model_name_prefix = glue("MACE_sa_{cohort}"), save.fig = FALSE)
  run_cox_models(df = df_sa, time = time, status = status, status_code = 2,
model_name_prefix = glue("Non_CV_Death_sa_{cohort}"), save.fig = FALSE)
}
```

```r
source("Code/cif.R") #code for obtaining CIF plots (unadjusted and adjusted)
model <- "primary"
source("Code/sdHR.R") #code for running subdistribution (Fine and Gray) model
if (only_injury == FALSE) { source("Code/sdHR-sa.R")} #for sensitivity analysis
removing non myocardial injury group and running subdistribution model
model <- "secondary"
source("Code/sdHR.R") #code for running subdistribution (Fine and Gray) model
if (only_injury == FALSE) { source("Code/sdHR-sa.R")} #for sensitivity analysis
removing non myocardial injury group and running subdistribution model
```

Once you have run the above code, go back to **fu_time <- c("1y", "5y")[1]**
Change this it to **fu_time <- c("1y", "5y")[2]** and rerun the code again.
This will produce results from 5y analysis.

Finally, if you have heart failure data, then run the last chunk of code. Remember that the HF_available object is specified in the "prepare-data-ipdma.R" file. **Only run this code once you are done with the code above (1y and 5y)**

You can run this code independent of the above. Just need to have sourced:
 - "prepare-data-ipdma.R" file
 - "functions-all-ipdma.R" file
 and define date and cohort objects

```r
if (HF_available == TRUE) {
 fu_time <- c("1y", "5y")[1] #define FU time (select 1y with [1] or 5y with [5])
  model <- "primary" #For this secondary analysis we will only use the primary model
  source("Code/HF-analysis-ipdma.R") #time and event vars for HF endpoint are
defined here
} else if (HF_available == FALSE) { print("you have no HF data in your cohort")}
```

If you have any queries regarding the code, please reply to Pedro and Andrew by email and we can advise or arrange a call.

Pedro Lopez-Alaya
Andrew Chapman

10.06.2025