**TO RUN FROM SCRATCH**

Files required:

* **data-provided**
  + OP\_prescriptions.csv
  + OP\_measures
    - environmental\_inhalers.csv
    - icsdose.csv
    - saba.csv
* 00.0\_source\_prescribing\_data.Rmd
* 00.1\_get\_OP\_codes\_from\_file.R
* 00.1\_get\_OP\_codes\_from\_scrape.R
* 00.2\_get\_NHSBSA\_codes.R
* 00.3\_get\_BNF\_info\_from\_OP.R
* 00.4\_get\_OP\_prescriptions.R
* 01.0\_source\_practice\_and\_demographic\_data.Rmd
* 01.1\_get\_ONS\_CCG\_shapes.R
* 01.2\_get\_NHS\_CCG\_patient\_numbers.R
* 01.3\_get\_QOF\_data.R
* 01.4\_get\_LSOA\_patient\_data.R
* 01.5\_get\_LSOA\_IMD\_data.R
* 02.0\_calculate\_prescribing\_measures.Rmd
* 03.0\_generate\_plots.Rmd

There are four R Markdown (.Rmd) files which call the other R scripts to run all of the operations. These are:

* 00.0\_source\_prescribing\_data.Rmd
* 01.0\_source\_practice\_and\_demographic\_data.Rmd
* 02.0\_calculate\_prescribing\_measures.Rmd
* 03.0\_generate\_plots.Rmd

Run the .Rmd files in order to run the project from start to finish.

00.0\_source\_prescribing\_data.Rmd scrapes the OpenPrescribing API, which could take > 18 hrs. We have therefore made available the original data we downloaded: ‘OP\_prescriptions.csv’, in the ‘data-provided’ folder (AVAILABLE ON REQUEST/HERE). Copy OP\_prescriptions.csv into a folder named data and run all chunks above ‘## 3. Collect Prescribing data from Open Prescribing.net’ in 00.0\_source\_prescribing\_data.Rmd to avoid this step. NOTE: be sure to also run ‘2.4. Call OpenPrescribing.net API to get drug details (name, is generic) from code’.

The ‘OP\_measures’ folder is also provided – this contains the BNF codes from the OpenPrescribing website that were previously scraped from the internet. The interface has since changed, so we provide the codes in .csv format.

Questions:

* Should functions be imported in a separate document or defined in the scripts they are used in?
  + Separate document when tidied repo up
* Do downloaded data files need to be saved as .csv?
  + Everytime URL is called download source data and store as well as output
* Should scripts that aim to return a variable be better as functions?
* How to define folders, e.g., ‘./data/’ – define this once in master script or check in every individual script?
  + Add .data and data-source creation to helper-function script
* Above affects whether each script should run in isolation

Check paper – happy with how its presented?

Aiming for folder structure:

./data – to contain all data. Files required for processing to be kept here

./data-source – to contain files used to generate data files required for processing

**Style guides**

<https://google.github.io/styleguide/Rguide.html>

<https://style.tidyverse.org/>

<https://style.tidyverse.org/files.html>

|  |  |
| --- | --- |
| **New file** | **Previous file** |
| 00.0\_source\_prescribing\_data.Rmd | OP\_setup.Rmd |
| 00.1\_get\_OP\_codes.R | get\_OP\_inhaler\_codes\_all\_measures.R |
| 00.2\_get\_NHSBSA\_codes.R | get\_OP\_inhaler\_codes\_all\_measures.R |
| 00.3\_get\_BNF\_info\_from\_OP.R | get\_OP\_inhaler\_codes\_all\_measures.R |
| OP\_codes.csv | environmental\_inhalers.csv |
| OP\_codes.csv | icsdose.csv |
| OP\_codes.csv | saba.csv |
| BNF\_codes\_measures | OP\_codes\_all\_measures |
| OP\_prescriptions | OP\_prescriptions\_all\_measures |
| 00.4\_get\_OP\_prescriptions.R | get\_OP\_prescriptions\_all\_measures.R |
| 01.0\_source\_practice\_and\_demographic\_data.Rmd | CCG\_all.Rmd |
| 01.1\_get\_ONS\_CCG\_shapes.R | CCG\_all.Rmd |
| 01.2\_get\_NHS\_CCG\_patient\_numbers.R | CCG\_all.Rmd |
| 01.3\_get\_QOF\_data.R | QOF\_2.R |
| 01.4\_get\_LSOA\_patient\_data.R | Make\_OP\_patients.Rmd |
| 02.0\_calculate\_prescribing\_measures.Rmd | Paper\_1\_export\_measures.Rmd |
| dates\_practices.csv | practices\_dates.csv |
| dates\_lsoa.csv | lsoa\_dates.csv |
| dates\_ practices\_ccgs.csv | practices\_ccgs\_dates.csv |

\*D:\LIDA\_ from RB\Inhalers\_COVID\_deprivation - CODE BY CODE\WORKNG\export\_measures - TEST\Rcode

**00.3\_get\_BNF\_info\_from\_OP.R**

Convert to use same log/comment system as 00.4\_get\_OP\_prescriptions.R

**00.4\_get\_OP\_prescriptions.R**

Should it print output to console as well as to log file? Maybe log is enough.

**CCG\_all.Rmd**

Only need the list of CCG ids and names but CCG\_all.Rmd downloads shape files etc too.

CHECK CHUNK NEAR THE BOTTOM – NHS BIT FOR CODE DOWNLOADS NO SHAPE

**Next step is to run Paper\_1\_export\_measures.Rmd.**

**Requires:**

CCG\_all.Rmd

Make\_OP\_patients.Rmd

qof\_reviews\_only.xlsx

* QOF\_2.R
* QOF\_check.R
* QOF\_lsoa.R

IMD – eatyourpeas

<https://github.com/eatyourpeas/imd>

Add functions to helper-functions.

**02.0\_export\_measures.Rmd – NOTES**

Can lose everything from line: 1393

## Check measures against those from OP

THERE IS A PROBLEM WHEN OBTAINING THE CODES – details below.

00.1\_get\_OP\_codes.R

- This code scrapes lists of BNF codes from the OpenPrescribing website from the URLs listedat line 68.

- However, OpenPrescribing recently updated these websites to use database function calls instead so no codes are imported

BNF\_codes\_from\_OP-required as prerequisite – This folder has been added. Should be included in repo as file to be included and 00.0\_source\_prescribing\_data.Rmd – should be adjusted.

We must therefore use data previously collected from the file:

Rcode > data > OP\_prescriptions\_all\_measures.csv

**Add plots folder**

03.0\_generate\_plots.Rmd