**README FOR CHAPTER 6 CODE**

Table of Contents

[General information 1](#_Toc36475204)

[Packages 2](#_Toc36475205)

[Data and figure information 2](#_Toc36475206)

[Program information 4](#_Toc36475207)

**DISCLAIMER:** Only population level data is required to run the simulation, therefore this has been provided and results of the simulation (used in Chapter 6) can be re-produced by running the provided code. Only code from section 3 and 4 is required to be ran to reproduce results.

# General information

* All programs run from a root/parent directory, where all the provided files and directories should be stored. At the top of each R program, set the working directory to this, or place an Rproject in this root directory and then nothing needs to be added to the code.
* The programs in section 1 and 2 are setting up the simulation. Discontinuation and restarting rates are derived from the statin users cohort, baseline CVD hazard is calculated, probability of non-cardiovascular death, and the function to run the simulation are defined. I am not able to share the patient level data used to derive these things, but the programs are provided for transparency.
* The programs in section 3 run the simulation, and the programs in section 4 produce results and figures. The population level metrics (such as discontinuation rates) that the simulations are ran on are all provided, as well as all required functions, so the simulation can be replicated and results presented in the paper can be re-produced using the code in sections 3 and 4. There is more details on the data provided in order to run the simulation in the ‘Data and figure information’ section.
* There is a batch file (AA\_run\_all\_batch.sh) that will run all the programs in order to carry out the analysis from the UNIX command line. However the working directory must be set at the top of each .R file in order to do this. If of interest, I would recommend running just a few of the programs (each program defined by treatment effect, and the statin cohort used), as each takes approximately 12 hours to run. There are 24 in total.
* Program 5.0 (in section 3) can be ran to produce figures of the extrapolated discontinuation rates that are used in the simulation (matches A6.1 Supplementary Figures 7, 8 and 9). These figures are also already provided in the /figures/ directory.
* The figures that will be produced when running the code for the main scenarios (treatment effect = 0.7, for the female and male cohorts) are already provided in the /figures/ directory.
* Programs were run using R.3.4.2.

# Packages

The package versions in the below table were used. These were the package versions installed on the remote computer I used for the analyses so some are quite old, however in most cases I see no reasons why the latest package versions wouldn’t work.

|  |  |
| --- | --- |
| Package | Version |
| foreach | 1.4.4 |
| doParallel | 1.0.11 |
| tidyverse | 1.3.0 |
| ggpubr | 0.1.8 |
| knitr | 1.20 |
| survival | 2.42-3 |

# Data and figure information

I provide information about the .RData files and .csv files that are provided, on which the simulations can be run. All .RData files are provided in */R\_out\_C6/* directory. All .csv files are provided in the */data/* directory. All figures (.jpeg) are provided in the */figures/* directory.

**DATA**

**1) generate\_model.agetime\_female/male.RData**: This contains the baseline hazard from the cox model (age as time scale), which is used to calculate the transition probabilities over the course of the simulation, for someone with a given 10 year risk. The function that does this is defined in the functions .RData file.

**2) generate\_risk\_range\_for\_each\_age\_agetime\_female/male.RData**: This contains the range of 10 year risks of patients of a given age that belong to the primary prevention cohort. I.e. a sensible range of 10 year risks to consider in scenarios of a given age.

**3) simulation\_functions\_agetime\_models\_female/male.RData:** This contains all the functions required to run the simulation. To understand what these functions do, A6.1 is essential reading.

*get.marg.risk.nodeath* > calculates marginal risk over a time period, for which the conditional risks each year are given

*get.10y.risk.nodeath* > For a given hazard ratio and age, it calculates the 10 year risk of someone with those properties

*get.HR* > for a given age and 10 year risk, it finds the corresponding hazard ratio

*get.cond.lifetime.risks* > for a given 10 year risk and age, calculates the conditional risk in each year of follow up

*run.simulation.agetime* > calculates the number of events expected over the course of the simulation, for a patient of a given age, 10 year risk and discontinuation rate (discontinuation rate must be yearly, as opposed to daily discontinuations, and therefore is not used in main simulation, it’s used to calculate risk assuming no discontinuation)

*run.simulation.agetime.CPRDdisc.cont.multiple.treat* > calculates the number of events expected over the course of the simulation, for a patient of a given age, 10 year risk and discontinuation rate, where the discontinuation rate takes the form of discontinuation rates we have derived from CPRD (used in main simulation)

**4) statin\_discontinuation\_rates\_for\_sim\_continuous\_multiple\_treatment\_periods\_agemod.RData**

Contains the survival curves used in the main simulation (first discontinuation rate stratified by age, second and third discontinuation, first and second restarting).

**5) statin\_discontinuation\_rates\_for\_sim\_continuous\_multiple\_treatment\_periods\_no1pres\_agemod.RData**

Contains the survival curves used in the sensitivity simulation where the statin cohort has had treatment periods of length one removed (first discontinuation rate stratified by age, second and third discontinuation, first and second restarting).

**6) statin\_discontinuation\_rates\_for\_sim\_continuous\_multiple\_treatment\_periods\_rates\_agemod.RData**

Contains the survival curves used in the main simulation where we adjust the discontinuation rates to be 5/6th, 2/3rd, and 1/2 of rates derived from CPRD.

**7) statin\_discontinuation\_rates\_for\_sim\_continuous\_multiple\_treatment\_periods\_rates\_no1pres\_agemod.RData**

Contains the survival curves used in the sensitivity simulation where we adjust the discontinuation rates to be 5/6th, 2/3rd, and 1/2 of rates derived from CPRD, but the statin cohort has had treatment periods of length one removed

**8) noncvd death rates riskcohort CPRD female/male.csv**

Contains the rate of non-cardiovascular death for females and males respectively. The ‘hazard’ column contains the probability of death for someone of that age, in that year.

**FIGURES**

**discontinuation1.extrapolation.jpeg:** Extrapolated discontinuation rate/survival curve for the first treatment period, stratified by age.

**discontinuation1.extrapolation\_no1pres.jpeg:** Same as previous figure but for cohort of statin users where treatment periods of length one were removed.

**discontinuation23.extrapolation.jpeg:** Extrapolated discontinuation rate/survival curve for the second and third treatment period.

**discontinuation23.extrapolation\_no1pres.jpeg:** Same as previous figure but for cohort of statin users where treatment periods of length one were removed.

**restarting extrapolation.jpeg:** Extrapolated restarting rate/survival curve for the first and second period of being off treatment.

**restarting\_extrapolation\_no1pres.jpeg:** Same as previous figure but for cohort of statin users where treatment periods of length one were removed.

**Figure3.jpg:** Figure 3 from the manuscript.

**Figure4.jpg:** Figure 4 from the manuscript.

# Program information

The table below details what is done in each program.

|  |  |
| --- | --- |
| Program | Definition |
| Section 1, statin cohort | |
| 0.1 | Install packages |
| 1.1 | Derive discontinuation and restarting rates for each treatment and restarting period |
| 1.2 | Derive discontinuation and restarting rates for each treatment and restarting period, where the first discontinuation period is stratified by age |
| 1.3 | Derive discontinuation and restarting rates for each treatment and restarting period, for the cohort where treatment periods of length one statin are removed |
| 1.4 | Derive discontinuation and restarting rates for each treatment and restarting period, where the first discontinuation period is stratified by age, for the cohort where treatment periods of length one statin are removed |
| 2.1 | Plot the survival curves from 1.1 on one graph (Figure 2, separate for discontinuation and restarting) |
| 2.2 | Plot the survival curves from 1.3 on one graph (separate for discontinuation and restarting) |
| 2.3 | Compare the age stratified discontinuation rates derived by the cox model (1.2), with Kaplan meier plots stratified by age bands |
| 3.1 | Baseline table |
| Section 2, set up simulation | |
| 1.1 | Fit cox model with age as time scale to the CVD primary prevention cohort, used to derive baseline hazard used in the simulation (female) |
| 1.2 | Fit cox model with age as time scale to the CVD primary prevention cohort, with 400,000 patients removed for calibration (female) |
| 1.3 | Fit cox model with age as time scale to the CVD primary prevention cohort, used to derive baseline hazard used in the simulation (male) |
| 1.4 | Fit cox model with age as time scale to the CVD primary prevention cohort, with 400,000 patients removed for calibration (male) |
| 2.1 | Define functions to run the simulation (female) |
| 2.2 | Define functions to run the simulation (male) |
| 3.1 | Generate range of risks for patients of specific ages (female/male) |
| 3.2 | Derive non-cardiovascular death rate at each age (female/male) |
| 4.1 | Extrapolate the discontinuation rates for use in the simulations (standard cohort and cohort with treatment periods of length one removed) |
| 4.2 | Adjust the extrapolated discontinuation rates to be 5/6th, 2/3rd and 1/2 of the rates derived from CPRD (standard cohort and cohort with treatment periods of length one removed) |
| Section 3, run simulation | |
| 5.0 | Plot the extrapolated discontinuation rates used in the simulation and save to /figures/ directory |
| 5 | Each program with prefix 5 runs the simulation for a group of scenarios. The discontinuation rate is that derived from CPRD. The treatment effect (Relative rate) and the statin cohort used (no 1 prescriptions, or standard cohort) used is given in the program name.  For example, *runsim\_female\_RR07* assumes a treatment effect of 0.7, and uses the standard statin cohort. However *runsim\_female\_RR07\_no1pres* assumes a treatment effect of 0.7, and uses the statin cohort where treatment periods of length one are removed.  Within each program, ages 40, 50 and 60 are considered, a suitable range of 10 year risk scores, and statin initiation at every possible year in follow up. |
| 6 | Same as description of p5 applies, however we also adjust the CPRD derived discontinuation rates to be 5/6th, 2/3rd and 1/2 of what they originally were.  Note that each program (from p5 and p6) corresponds to either Figure 3, Figure 4, or one of the 22 supplementary figures in Appendix 6.2. |
| Section 4, analysis and figures | |
| 1 | Each program with prefix p1 relates to a program with prefix p5 from section 3 and produces the appropriate figure. |
| 2 | Each program with prefix p2 relates to a program with prefix p6 from section 3 and produces the appropriate figure.  Note that each program (from p1 and p2) corresponds to either Figure 3, Figure 4, or one of the 22 supplementary figures in Appendix 6.2. |