**README FOR CHAPTER 5 CODE**

Contents

[General information 1](#_Toc36227781)

[Program information 1](#_Toc36227782)

**DISCLAIMER:** Simulated datasets to run this code on have not been provided. The code is provided for transparency.

# 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.
* I was not able to provide simulated raw data to run with this code, therefore the majority of the code is provided for transparency (I cannot provide the real data contractually). It was not straightforward to produce data that has the correct dependencies between the different time periods, given the real raw data was 10 different datasets representing the 10 time points. The corresponding author would be happy to help anyone with using this code upon request.
* The code for standard cox analyses is fairly straight forward (one stochastic imputation of data at baseline, then fitting a range of cox models).
* How to implement a MSM is covered extensively in the literature from the principal of mathematical/statistical principals,(1–3) and then also from a more pragmatic point of view.(4) The latter, details how data should be set up and how to calculate weights using the ipw package. Combined with the code provided by myself, this should be sufficient for understanding the approach taken and exactly what I have done.

# Program information

The table below details what is done in each program.

|  |  |
| --- | --- |
| **Standard cox analyses** | |
| 1.1 | Impute the data at baseline. |
| 1.2 | Re-formatting outcome variables. |
| 2.1 | Test for fractional polynomials of age, SBP and BMI. |
| 2.2 | Add these fractional polynomials to dataset. |
| 2.3 | Check for fractional polynomial of calendar time, using above fractional polynomials. |
| 2.4 | Add these fractional polynomials to dataset. |
| 3.1 | Develop cox model, check calibration of development and validation cohorts, and compare risks of development and validation cohorts. |
| 3.2 | Develop cox model with adjustment for calendar time, check calibration of development and validation datasets. |
| 3.3 | Create baseline table |
| **MSM analyses** | |
| 0.1 – 0.10 | Take datasets for each time period and create outcome data for interval censored data, for the respective interval. For example, in the first time period, followup begins at 0 and ends at 183 days. For the second time period, follow up starts at 183 days and ends at 365. Outcome indicator only = 1 when events is in a particular time period. |
| 1.1 | Combine 10 baseline datasets for each time period into one dataset that can be used to model interval censored data. Dataset includes baseline variables and time varying variables, and interval censored outcome data (of which was calculated in 0.1 – 0.10).  Within each patient, run a last observation carried forward (locf) imputation on the time varying variables. Then run a next observation carried backwards (nocb) to impute any data missing prior to a non-missing value. |
| 1.2 | Run one stochastic imputation (one multiple imputation by chained equations) on the baseline data. |
| 1.3 | Assign this imputed baseline data to the time varying variables at baseline also, if it is still missing after program 1.1 (some variables with missing values at each time period will still have missing data). Then use locf to impute the time varying variables at each point in time with this stochastically imputed value.  There is now no missing data. Whenever possible, we have used locf or nocb imputation to maintain consistency in variables within a patient across time points. We have only used stochastic imputation to impute a value if a patient had no data across every time point. We use the one stochastically imputed value across every time point. |
| 2.1 | Fit the interval censored cox and MSM models. |
| 2.2 | Compare risks of validation cohort before and after calendar time was introduced, for the interval censored cox and MSM models. |
| 2.3 | Calibration plots for the MSM and interval censored cox model |

REFERENCES

1. Robins J, Hernan M, Brumback B. Marginal structural models and causal inference in epidemiology. 2000;

2. Hernán M, Brumback B, Robins J. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. Epidemiology. 2000;11(5):561–70.

3. Sperrin M, Martin GP, Pate A, Van Staa T, Peek N, Buchan I. Using marginal structural models to adjust for treatment drop-in when developing clinical prediction models. Stat Med. 2018;37(28):4142–54.

4. Wal WM van der, Geskus RB. ipw: An R Package for Inverse Probability Weighting. J Stat Softw. 2011;43(13).