Propensity score weighting to estimate a marginal effect

1) Load the database dataCOHORT. (https://github.com/chupverse/gcomputation/tree/main)

```
library(gcomputation)
data(dataCOHORT)
```

2) Looking at the descriptive table, check the positivity assumption[1].
   You will find the descriptive table in the "description_dataCOHORT" document (https://github.com/chupverse/causal-workshop/tree/main/session2).
   For the remainder of the practical assignment, we will choose to exclude patients under the age of 30.

3) Compare the outcome (VAP) observed in each group.

4) Looking at the DAG, choose the prognostic factors to adjust on.
   You will find the code of the DAG in the "dag_dataCOHORT" document (https://github.com/chupverse/causal-workshop/tree/main/session2). Copy and paste it on Dagitty in the section *Model code*.
   For the remainder of the practical assignment, we will choose to adjust on the age, the BMI, the gender, the Glasgow Coma Scale, the injury and the PAO2/FIO2.

5) Construct the exposure model using multivariate logistic regression.
   (You have to work in complete-case.)

6) For each subject, calculate his propensity score, i.e. the probability of being treated according to the individual characteristics.

7) Check the positivity assumption by plotting the distribution of the propensity score in the two separate groups.

8) Calculate weights.

The individual weights for a subject *i* is defined by[2]:

$$w_i = \frac{A_i Pr(A_i = 1)}{\pi_i} + \frac{(1 - A_i) Pr(A_i = 0)}{1 - \pi_i}$$

Where π_i is the propensity score for subject *i*, and A_i is equal to 1 if the subject *i* is treated and 0 if not treated.

9) Check the correct specifications for your model by calculating the standardized mean difference[3] (SMD) of the covariates AGE and SEX in the counterfactual population.

For continuous covariates, SMD is defined by:

$$d = \frac{\bar{z}_{A=1} - \bar{z}_{A=0}}{\sqrt{\frac{s^2_{A=1} + s^2_{A=0}}{2}}}$$

Where $\bar{z}_{A=1}$ and $\bar{z}_{A=0}$ are the weighted means of the characteristic of interest in the population treated and untreated, respectively. $s^2_{A=1}$ and $s^2_{A=0}$ are the weighted variance of the characteristic of interest in the population treated and untreated, respectively.

For categorical covariate, SMD is defined by:

$$d = \frac{\hat{p}_{A=1} - \hat{p}_{A=0}}{\sqrt{\frac{\hat{p}_{A=1}(1-\hat{p}_{A=1}) + \hat{p}_{A=0}(1-\hat{p}_{A=0})}{2}}}$$

Where $\hat{p}_{A=1}$ and $\hat{p}_{A=0}$ are the proportion of patients presenting the characteristic of interest in the weighted population among those treated and untreated, respectively.

10) Calculate the average treatment effect on the entire population (ATE) by computing the difference of proportion of VAP between treated and untreated. (binary outcome)

Bonus questions

11) Calculate the marginal effect of the group on the death (censored outcome).
12) Estimate the confidence intervals of the difference of VAP between treated and untreated by bootstrapping the previous step.
13) Use the Multiple Imputation by Chained Equations (MICE) algorithm – or any other imputation method – to avoid excluding patients with missing data. To combine bootstrap estimation with multiple imputation, we choose to bootstrap the completed data sets[4].
14) It has been suggested that consideration of risk factors (causes of the outcome) and confounding factors (causes of both the treatment and the outcome) in the propensity scores reduces bias and variance[5]. Use the Lasso method to select the covariate related with the VAP[6].

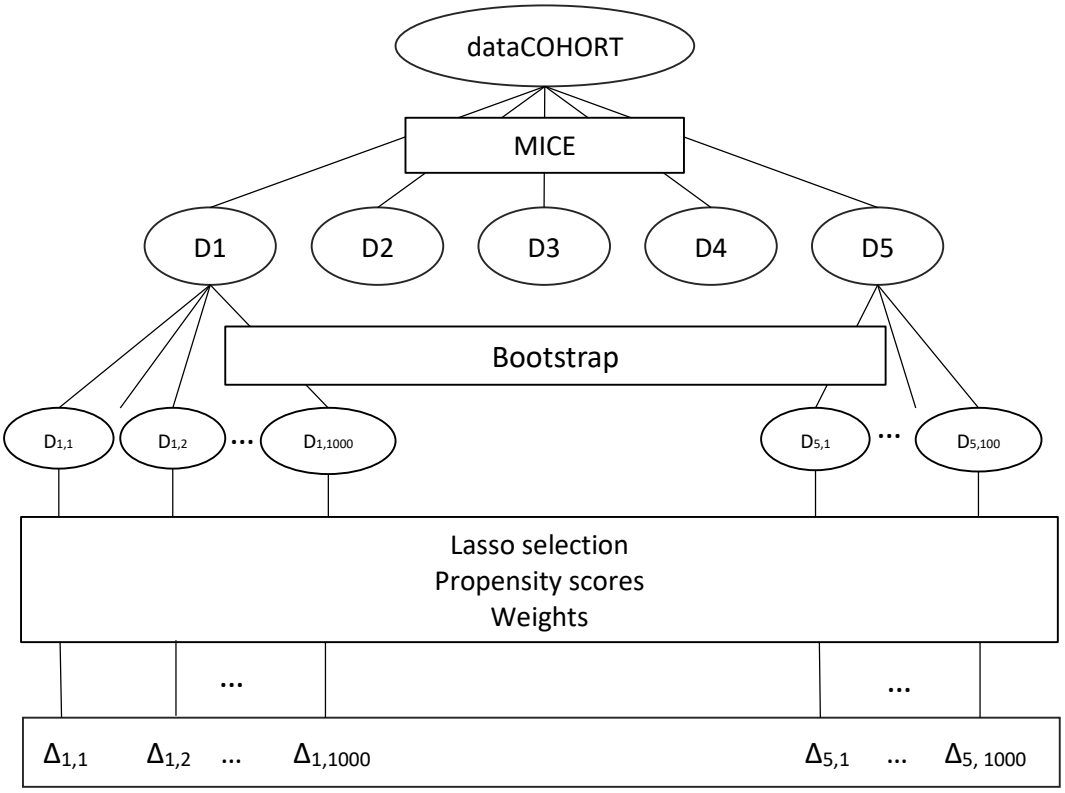

Figure 1 - Summary diagram of the chosen analysis methodology

References

(1) Léger, M.; Chatton, A.; Borgne, F. L.; Pirracchio, R.; Lasocki, S.; Foucher, Y. Causal Inference in Case of Near-Violation of Positivity: Comparison of Methods. *Biometrical Journal* **2022**, *64* (8), 1389–1403. https://doi.org/10.1002/bimj.202000323.

(2) Austin, P. C.; Stuart, E. A. Moving towards Best Practice When Using Inverse Probability of Treatment Weighting (IPTW) Using the Propensity Score to Estimate Causal Treatment Effects in Observational Studies. *Stat Med* **2015**, *34* (28), 3661–3679. https://doi.org/10.1002/sim.6607.

(3) Austin, P. C. Using the Standardized Difference to Compare the Prevalence of a Binary Variable Between Two Groups in Observational Research. *Communications in Statistics - Simulation and Computation* **2009**. https://doi.org/10.1080/03610910902859574.

(4) Schomaker, M.; Heumann, C. Bootstrap Inference When Using Multiple Imputation. *Stat Med* **2018**, *37* (14), 2252–2266. https://doi.org/10.1002/sim.7654.

(5) Chatton, A.; Le Borgne, F.; Leyrat, C.; Gillaizeau, F.; Rousseau, C.; Barbin, L.; Laplaud, D.; Léger, M.; Giraudeau, B.; Foucher, Y. G-Computation, Propensity Score-Based Methods, and Targeted Maximum Likelihood Estimator for Causal Inference with Different Covariates Sets: A Comparative Simulation Study. *Sci Rep* **2020**, *10* (1), 9219. https://doi.org/10.1038/s41598-020-65917-x.

(6) Shortreed, S. M.; Ertefaie, A. Outcome-Adaptive Lasso: Variable Selection for Causal Inference. *Biometrics* **2017**, *73* (4), 1111–1122. https://doi.org/10.1111/biom.12679.