Item S1. Inverse Probability of Treatment Weighting and Propensity Scores Matching Approaches

1. *Inverse Probability of Treatment Weighting Approach*

The use of the weights is a method to simulate what would have happened had everyone in the population experience both levels of exposure (Hernan and Robins JECH 2006). A simple application of the weights creates a pseudo-population of double size but this double size does not imply higher precision because each individual is used twice. By applying stabilization (multiplying the weight by the probability of being exposed for those exposed and the probability of being unexposed for those unexposed), we are able to create a pseudo-population with similar percentage of patients exposed in each level of the covariates as the overall percentage in the study population. As explained in Hernan and Robins 2006, inverse probability of treatment weighting (IPTW) mimics a design I randomized experiment that eliminates selection bias by observed characteristics between warfarin users and non-users and with use of stabilized weights the pseudo-population has the same size as the original population.

Implementation

The following steps show how to implement IPTW:

Step 1. Run a multivariate logistic regression model to predict the probability of a patient being exposed ($\hat{p}_{Ti}$), adjusted for all variables listed in Table 1.

Step 2. Compute a weight for each patient in each exposure group:

For a patient *i* in the exposed group with N_T patients, $w_i = \frac{1}{\hat{p}_i}$

For a patient *j* in the unexposed group with N_C patients, $w_j = \frac{1}{1-\hat{p}_j}$

Step 3. Compute the Stabilized Weights (sw)

To compute stabilized weights[1,2] one multiplies the weight found in Step 2 above by the probability of being exposed (p_T) for those exposed and the probability of being unexposed for those unexposed $(1 - p_T)$. More specifically the stabilized weight, sw, is defined as:

For a patient *i* in the exposed group, $sw_i = p_T * w_i$

For a patient j in the unexposed group, $sw_j = (1 - p_T) * w_i$

In our sample, this would imply multiplying by 0.15 for warfarin users and 0.85 for warfarin non-users which are the percentages of patients in the treated and not treated groups, respectively.

Step 4. If necessary, truncate weights to the value 10 (0.1) if stabilized weights are found to be too large (small). (Harder et al 2010)

As a check and also as a tool to decide on a model specification, the estimated weights are required to have a mean of 1 (approximately). (Cole and Hernan AJE 2008)

Inverse Probability of Censoring Weighting approach

We follow the same steps (1-4) to compute stabilized censoring weights. In this case, in Step 1 one starts with a logistic regression model to predict the probability of a patient being censored ($\hat{p}_{Ci}$), adjusted for all variables of interest and stabilized weights are computed using the proportion of patients that are censored.

The final weights

Final weights are computed by multiplying the stabilized and trimmed weights for exposure with the stabilized and trimmed weights for censoring. If necessary, further trimming is applied. Because censored differed by outcome, separate weights are compute for each outcome separately.

Analysis

Comparison of statistics (mean, proportions) between exposed and unexposed in the pseudo-population can be done through the use of procedures (functions) that take into accounts weighting of observations. Similarly standardized differences were computed with the use of the weighted means, weighted proportions and weighted standard deviations.

We used weighted robust Cox proportional hazards models to obtain parameter estimates and SE.

2. *Propensity Scores Matching Approach*

The propensity score is the adjusted probability of a patient being a warfarin user. (cite Rosenbaum and Rubin 1983) This probability is computed using the same multivariate logistic regression model run in Step 1 of the IPTW analysis. One-on-one matching was conducted using the psmatch2 program in STATA 12 where warfarin users are matched with non-users with a difference in propensity scores of no greater than 0.01. (Leuven and Sianesi 2013)

Works Cited

Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology 2000; 11:550-560

Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. Am J Epidemiol. 2008; 168(6):656–664.

Harder , V. S. Stuart , E. A. Anthony , J. C. Propensity score techniques and the assessment of measured covariate balance to test causal associations in psychological research. Psychological Methods 2010; 15:234-249. doi:10.1037/a0019623

Rosenbaum, PR and Rubin, DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983; 70(1): p. 41-55.

Leuven E and Sianesi B: PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. Available at http://ideas.repec.org/c/boc/bocode/s432001.html. (accessed 1/10/2014), 2003