INITIATORS SAS MACRO

Authors: Roger Logan, Goodarz Danaei, Miguel Hernán

The INITIATORS macro is designed to analyze observational longitudinal data to estimate the effect of interventions sustained over time. The premise is to emulate the design and analysis of a hypothetical randomized trial so the Macro uses the language of randomized clinical trials applied to an observational setting. The Macro is capable of conducting the observational analogs of the intention-to-treat, per-protocol and as-treated analyses (see below). All analyses are conducted using pooled logistic regression to approximate the hazard ratio from a proportional hazard Cox model. For more details see the article by Danaei et al cited at the end of this document.

Data Structure

- Please see the appendix of the cited paper for some technical detail on the structure of the data and how it is used in the Macro.

- The data should be in a person-time format so the id and period variables are required for the analysis.

- It would save time if the user sorts the data by id and period before running the Macro. The Macro will sort the data by id and period if the option *sorted=0* is used.

- Periods should be consecutive with no missing periods for any participant.

- Periods should start at 0. If the first period is not 0 then the user should set the option *baseline_offset=X* to indicate the value of the first period (here *X*).

- The first period of each person should be eligible (however eligibility is defined by the user).

- There should be no missing values for treatment.

Analyses

1. Intention-to-treat analysis

- To mimic a proportional hazard Cox model comparing initiators with non-initiators of treatment the user should set the options to:

use_weights=0, use_censor=0, model_var=_assigned_treatment

- The output of this analysis is a hazard ratio comparing initiators to non-initiators of treatment.

2. Per-protocol analysis

- If the option *use_censor=1*, the Macro will censor person-times once a person-trial stops taking the initial treatment value (i.e., treatment value at the baseline of that trial).

- The default treatment variable for the outcome model will be the initial treatment value.

- If the user sets the option *include_regime_length=1*, a new variable named _time_*on_regime* is added to the database. This variable stores the duration of time that the patient has been on the current treatment value. This variables and its squared term will be added to the models for denominator of the IP weight as predictors of future treatment.

- The output of this analysis is a hazard ratio comparing what would have happened had everyone adhered to their assigned treatment (i.e continuous treatment versus no treatement).

3. As-treated analysis

- If *use_censor=0*, the Macro will continue expanding the eligible person-trials until the end of follow-up or occurrence of an event or censoring. The default functional form for the treatment in the outcome model is cumulative treatment and its squared term but the user can define any other function of treatment by providing the functional form to the macro *%create_final_variables* and adding the name of that variable to *model_var=*. If the user chooses to redefine the functional form of the treatment, the new user-defined macro should be located after the *%include .../INITIATORS.sas* statement in the call.

- The elements most commonly used to define a function of treatment for the outcome model are included in a separate macro called *%building_blocks* and include the cumulative dose for each trial (*_dose_*), the cumulative dose for the entire history of treatment for a participant (*_cumA_*) and two temporary arrays one including treatment at each time (*_treat_*) and the other including cumulative treatment at each time (*_dosesum_*).

- The output of this analysis is a hazard ratio which will be defined by the specific functional form of treatment used in the disease model.

Other Options and Macro Parameters

- The variables to be included in the numerator of the IP weights models should be defined in *cov_switchn*, the *model_switchn* allows the user to include interaction terms between covariates and the *class_switchn* can be used to include categorical variables. The same applies to the denominator weights with the three options *cov_switchd, model_switchd* and *class_switchd*.

- The variables to be included in the outcome model should be defined in *outcomeCov_var*. Any interaction terms to be included can be defined using *outcomeCov* and any categorical variables in *outcomeClass*.

- The user can choose to include the trial's time and follow-up time in the outcome model(s) by setting *include_expansion_time_case=1* and *include_followup_time_case=1*. By default, both linear and squared terms will be added to the outcome models.

- If *use_weights=1*, the user can chose any combination of the four possible outcome models by setting the corresponding option to 0 or 1: no weights (*run_unweighted_analysis=1)*, original weights (*run_weighted_analysis=1)*, weights truncated at 1st and 99th percentile (based on distribution of weights in the entire sample) (*run_p99_analysis)*, or weights truncated with user-defined thresholds defined by options *lower_weight=* and *upper_weight=* (*run_user_limits_analysis=1)*. If no option is specified, the Macro will run the model with original weights only.

- The option *lag_p_nosw* is by default set to 1 to skip the first time point in each trial when IP weights are being calculated. The reason is that the final model is expected to have the baseline covariates in the model so there is no confounding by these covariates and by skipping the first factor in the IPW product, we gain some efficiency by reducing the number of factors being multiplied to calculate the weight at each time. The user can set this option to zero which will then increase the maximum and variance of IP weights.

- The two options *eligible_wt_0* and *eligible_wt_1* allow the user to exclude some observations from the IP weight models. For example, if it is assumed that a patient will stay on treatment for at least 2 months, the first 2 months after treatment initiation by definition have a probability of staying on treatment of 1.0 and should thus be excluded from the weight models for the treated. So, the user has to define a variable that indicates that these 2 observations are ineligible for the weight model for the treated and put the name of that variable in front of the *eligible_wt_1* option.

- The Macro includes a set of variables and options to use IP weights to adjust for informative censoring (e.g., due to loss to follow-up). These include *cense=* which indicates the censoring variable (1:censored, 0: not censored), *model_censen* and *class_censen* to indicate which variables to use in the models for numerator of the censoring weights, *model_censed* and *class_censd* for the model for denominator of the censoring weights and *pool_cense=* which indicates whether the models should be fit on the entire data (when set to 1) or be fit separately for those who were treated vs. untreated in the previous month (if set to 0).

- The user can set the option *final_analysis=0* when exploring the weight models and distribution of weights to save time by not running the logistic model for the outcome.

- The user can run a conditional outcome model by using the options *where_case* and *where_var*. In such cases, the marginal model may be skipped or not using the *run_base_model*.

- Setting *print_option=noprint* suppresses the print output from the weight and outcome models.

- If *check_missing=1* the Macro will run a proc means to check for the missing observations for all the covariates in the outcome model.

- The user can select only part of follow-up to be included in all analyses by setting *first_period=* and *last_period=* to the desired first and last periods (inclusive).

- The user may also limit the periods included in the outcome model by setting *first_followup=* and *last_followup=* to the desired first and last follow-up periods. Notice that periods not included in this range will still contribute to the weight models.

Notes

- Notice that if the baseline values of confounders for each person (when they first become eligible) has been used to stabilize the weights, which is usually the case, then the final model has to be adjusted for these covariates too to adjust for confounding.

- The option *use_censor* is related to artificial censoring due to switching treatment and should not be confused with the group of options that are related to censoring due to loss to follow-up. (The latter are: *cense*, *pool_cense, model_censen, model_censed, cov_censen, cov_censed, class_censen, class_censed*.)

- In the per-protocol analysis, when the Macro artificially censors person-time after the patient discontinues treatment, the Macro assigns a missing value for the event for all the last observations in all the person-"trials" that are created by expanding the data from that particular patient and start before treatment discontinuation. Therefore, these last observations will not contribute to the outcome model.

- If *calculate_var=1* all the coefficients and their standard errors (SE) from the outcome model have to be non-missing. So, if for any reason (e.g. a variable being a linear combination of other variables in the model) the coefficient or its SE is missing, the Macro will stop and return an error in matrix conformability.

CITATION

Danaei G, Garcıa Rodriguez LA, Cantero OF, Logan R, Hernán MA. Observational data for comparative effectiveness: an emulation of randomized trials to estimate the effect of statins on primary prevention of coronary heart disease. submitted.

FUNDING SOURCE