PSS weighted analysis macro- user guide (last update December 24, 2020)

Description and citation:

This macro performs propensity score (PS) adjusted analysis using stratification for cohort studies from an analytic file containing information on patient identifiers, exposure, confounding variables or pre-computed PS, and binary outcomes/censoring time. In the first step, patients from non-overlapping regions of PS distributions are trimmed. Next, PS strata are created using either the distribution of PS in the exposed group only or the entire cohort as specified by the user. Next, this macro calculates weights targeting the ATT (Average Treatment effect among the Treated) or the ATE (Average Treatment Effect in the whole population) as specified by the user. Finally, weighted generalized linear models or weighted Cox-proportional hazards model provides adjusted effect estimates along with confidence intervals calculated using robust variance estimates to account for weighting. If you use this macro, please cite the following publications- which contain further technical details on the weighting methods used in this macro following stratification.

> Desai RJ, Rothman KJ, Bateman BT, Hernandez-Diaz S, Huybrechts KF. A Propensity-score-based Fine Stratification Approach for Confounding Adjustment When Exposure Is Infrequent. *Epidemiology*. 2017 Mar;28(2):249-257.

> Desai RJ, Franklin JM. Alternative approaches for confounding adjustment in observational studies using weighting based on the propensity score: a primer for practitioners. *BMJ*. 2019 10 23; 367:l5657

Structure:

```
%macro fine_stratification (in_data= , exposure= , PS_provided= no ,
ps_var= , ps_class_var_list= , ps_cont_var_list= , interactions= ,
PSS_method=exposure,
estimand = ,
n_of_strata= , out_data= PS_FS, id_var= ,
effect_estimate= ,
outcome= , survival_time= , time_unit= ,
out_excel= , work_lib= clean);
```

Required parameters:

1. **In_data =** Input analytic file with patient identifiers and all their covariate, exposure, and outcome information.

2. **exposure =** treatment variable.

   NOTE- This variable has to be in binary numeric format where 0= unexposed (or reference), 1= exposed. This formatting is critical to build a reliable PS model

3. **ps_var=** name of the variable for the calculated PS (if PS_provided=no) or pre-computed PS (if PS_provided= yes)

4. **ps_class_var_list =** Discrete variables to be included in the PS model (include all binary/categorical variables here regardless of the format. This list goes in the 'Class' statement of 'proc logistic' for the PS model)

5. **ps_cont_var_list =** Continuous variables to be included in the PS model

   NOTE- Even if you are using a pre-computed PS, listing out variables included in the PS model under 'ps_class_var_list' and 'ps_cont_var_list' is important because this information is used to create covariate distribution Tables before and after PS weighting to assess whether weighting reduced imbalances.

6. **n_of_strata=** Number of PS strata to be created

7. **id_var:** Patient identification variable

8. **estimand:** Desired estimand
   - Available options
     - ATT (Average Treatment effect among the Treated- weighting of reference patients conducted based on the number of exposed in each stratum; all exposed receive weight of 1)

ATE (Average Treatment Effect in the whole population weighting of both exposed and reference patients conducted based on the number of total patients in each stratum)

9. **effect_estimate:** Desired effect estimate

        Available options for binary outcomes
                RD (risk differences using proc genmod link- identitiy)
                RR (risk ratios using proc genmod link- log)
                OR (odds ratios using proc genmod link- logit)
        Available options for time to event outcomes
                HR (Hazard ratios using proc phreg)

10. **outcome:** Binary outcome variable where 1=occurrence of the outcome

11. **survival_time:** Follow-up time variable indicating survival. Only required if effect_estimate=HR

12. **time_unit:** Unit of the provided survival_time variable. Only required if effect_estimate=HR. Used to calculated incidence rate and 95% confidence intervals per 100 person years

        Available options
                days
                years

13. **out_excel:** Path and name of the excel file to be saved which will contain all the outputs (For details of the contents of this file, please refer to <u>page 5</u> of this Guide)

Optional parameters:

1. **PS_provided:** Parameter specifying whether a PS model needs to be built or already provided. Default is (PS_provided= no), meaning PS needs to be calculated within the macro. If PS is already calculated, set (PS_provided= yes) and specify which variable in your dataset contains the PS under 'ps_var'.

2. **Interactions:** To include interactions among covariates while constructing the PS model (eg. age*gender, this goes in the model statement of proc logistic)

3. **PSS_method:** The method to be used for creating PS strata. The default is (PSS_method=exposure), meaning PS strata will be created using the distribution of PS only among the exposed group and unexposed will be assigned to these strata based on their PS. The other available option is (PSS_method=cohort), which uses the PS

distribution of the entire cohort to create strata.

4. **Out_data:** Name of the SAS dataset that contains information of strata and weights for each patient from the input dataset (this is the dataset on which the final weighted regression models are run)

5. **Work_lib:** To clean out the work library after the macro is executed (by default, it will do this; set work_lib=no (or any word other than 'clean') to turn this feature off).

Example dataset and calls:

Following calls can be directly executed in the 'ps_fs_binary_example' and 'ps_fs_survival_example' datasets that are provided with this guide.

```
** The following two macros need to be referenced first.

1, the stratification and analysis macro- %fine_stratification;

%include 'C:\Users\rjd48\Desktop\PSS weighted analysis\PSS weighted analysis.sas';

* 2, the macro for creating patient characteristic tables- %table1

NOTE- this macro is invoked by calling the %fine_stratification macro, so needs to be
references but no need to separately call it;

%include 'C:\Users\rjd48\Desktop\PSS weighted analysis\Weighted Table 1s.sas';

** binary outcome example;

%fine_stratification (in_data= pss.ps_fs_binary_example,
exposure= treatment, PS_provided= no , ps_var= ps,
ps_class_var_list= c1-c9 , ps_cont_var_list= c10,
interactions= c1*c2 c7*c9,
PSS_method=exposure,
estimand= att,
n_of_strata= 50, out_data= PS_FS, id_var= obs, effect_estimate= rr,
          outcome= outcome,
out_excel= C:\Users\rjd48\Desktop\PSS weighted analysis\example_results_RR);

** time to event outcome example;

%fine_stratification (in_data= pss.ps_fs_survival_example,
exposure= treatment, PS_provided= no , ps_var= ps,
ps_class_var_list= c7-c10 c1 c3-c5 , ps_cont_var_list= c9 c2,
interactions= c1*c2 c7*c9,
PSS_method=exposure,
estimand= att,
n_of_strata= 50, out_data= PS_FS, id_var= id, effect_estimate= hr,
outcome= outcome, survival_time= survt, time_unit= days,
out_excel= C:\Users\rjd48\Desktop\PSS weighted analysis\example_results_HR);
```

Diagnostics and results excel file output from the macro

Several diagnostics are produced by default in the output excel file (name as specified under 'out_excel' in the macro) along with results from models based on parameters specified above. These contents are in line with reporting practices noted in Desai & Franklin *BMJ*. 2019 10 23; 367:l5657. Following components in separate excel sheets are generated in the output file.

1. Non-overlap trimming summary: Number of observations trimmed in the exposed and reference groups due to non-overlapping PS distributions
2. Stratification and weighting summary: A large table showing total number of strata in each row, range of propensity scores within each stratum, total number of exposed and reference patients in each stratum, total number of events within each stratum, weights for each stratum, total number of observations and events after weighting within each stratum
3. Counts for the unweighted and weighted samples
4. Plots for PS distribution before and after weighting
5. Box plot for weight distribution by exposure group
6. Covariate balance plot showing standardized differences before and after weighting
7. Unweighted Table 1: Unweighted distribution of the covariates that are part of the propensity score models along with crude proportion or mean difference and standardized difference
8. Weighted Table 1: Distribution of the covariates that are part of the propensity score models after propensity score weighting along with crude proportion or mean difference and standardized difference
9. Overall balance: C-statistic for the original PS model and post-weighting C-statistic as a measure of the overall covariate balance achieved through PS-weighting (values closed to 0.5 post-weighting indicate better overall balance)
10. Risk of rate table: unweighted and weighted event counts and total number at risk and risk estimates (for binary outcomes) OR cumulative person-time and incidence rate estimates (for time-to-event outcomes) reported per 100 person-years with 95% CI based on Poisson distribution
11. Effect estimate: Results from the regression models before and after PS-weighting