

PSS weighted analysis macro- user guide

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. Finally, this macro conducts weighting of unexposed (or reference) subjects based on the number of exposed subjects residing in their PS-stratum and uses weighted generalized linear models or weighted Cox-proportional hazards model to provide adjusted effect estimates. If you use this macro, please cite the following publication- which contains further technical details on the weighting method 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*. 2016 Nov 29. [Epub ahead of print] PubMed PMID: 27922533.

PSS weighted analysis macro- user guide

Structure:

```
%macro fine_stratification (in_data= , exposure= , PS_provided= no ,  
ps_var= , ps_class_var_list= , ps_cont_var_list= ,  
PSS_method=exposure,  
n_of_strata= , out_data= PS_FS, id_var= ,  
effect_estimate= ,  
outcome= , survival_time= ,  
out_excel= ,  
ps_dist_plots= no, ps_dist_figure= );
```

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. **effect_estimate**: Desired effect estimate

Available options for binary outcomes

RD (risk differences using proc genmod link- identity)

RR (risk ratios using proc genmod link- log)

OR (odds ratios using proc genmod link- logit)

Available options for time to event outcomes

PSS weighted analysis macro- user guide

HR (Hazard ratios using proc phreg)

9. **outcome:** Binary outcome variable where 1=occurrence of the outcome
10. **survival_time:** Follow-up time variable indicating survival. Only required if effect_estimate=HR
11. **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 [page 5](#) 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. **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.
3. **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). The default is a dataset called 'ps_fs'- which will be saved in the work library.
4. **ps_dist_plots:** Parameter specifying whether plots for PS distribution are requested. The default is (ps_dist_plots= no), when set to (ps_dist_plots=yes) it will give plots before and after PS weighting.
5. **ps_dist_figure:** Path and name of the pdf file to be saved which will contain PS distribution plots. Only needed if (ps_dist_plots=yes).

PSS weighted analysis macro- user guide

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= ps_fs_binary_example, exposure= treatment,  
PS_provided= no , ps_var= ps, ps_class_var_list= c1-c9 , ps_cont_var_list=  
c10, PSS_method=exposure, n_of_strata= 50, out_data= PS_FS, id_var= obs,  
effect_estimate= rr, outcome= outcome, survival_time= ,  
out_excel= C:\Users\rjd48\Desktop\PSS weighted analysis\example_results,  
ps_dist_plots= yes,  
ps_dist_figure= C:\Users\rjd48\Desktop\PSS weighted  
analysis\example_ps_dist);
```

** time to event outcome example;

```
%fine_stratification (in_data= ps_fs_survival_example, exposure= treatment,  
PS_provided= no , ps_var= ps,  
ps_class_var_list= c1 c3-c5 c7-c10 , ps_cont_var_list= c2 c9,  
PSS_method=exposure,  
n_of_strata= 10, out_data= PS_FS, id_var= id, effect_estimate= hr,  
outcome= outcome, survival_time= survt,  
out_excel= C:\Users\rjd48\Desktop\PSS weighted analysis\example_results,  
ps_dist_plots= yes,  
ps_dist_figure= C:\Users\rjd48\Desktop\PSS weighted  
analysis\example_ps_dist);
```

PSS weighted analysis macro- user guide

Contents of the output files:

This macro will analyze the data based on the parameters specified above and provide a detailed output in an excel file (name as specified under 'out_excel' in the macro) containing the following components in separate excel sheets,

1. 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
2. Table 2: 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
3. Table 3: 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 closer to 0.5 post-weighting indicate better overall balance)
4. Table 4: 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)
5. Table 5: Results from the regression models before and after PS-weighting

In addition, if 'ps_dist_plots'= yes and a pathname is provided under 'ps_dist_figure', the macro also gives a pdf file containing propensity score curves for crude and weighted population along with vertical lines indicating the boundaries of the formed PS-strata