Process macro- user guide

Author: Kehang Li

Structure:

```
%macro process ( in_data_t= ,
                 in_data_m= ,
                 diag_dim= ,
                outcome_dim= ,
                fill_dt= ,
                DAYS_SUP= ,
                BRND_NM= ,
                fst_dt= ,
                lst_dt= ,
                patid= ,
                input_path_1= ,
                input_path_2= ,
                output_path= );
```

Required parameters:

1. **in_data_t:** Input analytic SAS treatment dataset name with patient identifiers that have purchase records of the specified pharmaceutical drug and all their covariate information.
2. **in_data_m:** Input analytic SAS dataset containing the patient identifiers, exposure, outcome, and any predefined covariates. Datasets will not be modified.
3. **diag_dim：** The number of diagnostic codes (ICD-9) variables dimension, parameter specified one or more times to indicate the ICD-9 records from various data source.
4. **outcome_dim:** The number of outcome diseases dimension, parameter specified one or more target outcomes in the study, for unique drug intake treatment.
5. **fill_dt:** The first occurrence date of a specified pharmaceutical drug purchase records.
6. **DAYS_SUP:** The duration between the first occurrence date and last occurrence date of a specified pharmaceutical drug purchase records, with time unit of day.
7. **BRND_NM:** The string variable that indicates the name of purchased pharmaceutical drug.
8. **fst_dt:** The first occurrence date of a specified disease medical records.
9. **lst_dt:** The first occurrence date of a specified disease medical records.
10. **patid:** Patient identification variable.
11. **input_path_1:** The path to a directory where the process macro can store temporary files. There should be enough space in the directory to hold a second copy of the input cohort and each of the dimensions.
12. **input_path_2:** The path to a directory where the input SAS treatment datasets are store. There should be enough space in the directory to hold a second copy of the input cohort and each of the dimensions.
13. **output_path:** The path to a directory where the process macro can store output datasets, tables and files.

Example dataset and calls:

```
** The following macro need to be referenced first.
1. the preproessing and analysis macro- %process;
%include '/scratch/yf31/uhc/kl/process.sas';

NOTE- this macro is invoked by calling the %fine_stratification macro, so needs to be
references but no need to separately call it;
%macro process ( in_data_t= phar_betablocker,
                 in_data_m= med_phar_dt,
                 diag_dim= 5,
                outcome_dim= 8,
                fill_dt= fil_dt,
                DAYS_SUP= sup_days,
                BRND_NM= BRND_NM,
                fst_dt= first_dt,
                lst_dt= last_dt,
                patid= id,
                input_path_1= /scratch/yf31/uhc,
                input_path_2= scratch/yf31/uhc/phar_dt,
                output_path= scratch/yf31/uhc/outcome );
```