PharmaSUG 2024 - Paper RW- 227

A SAS® Macro Approach: Defining Line of Therapy Using Real-World Data in Oncology

Yu Feng, Merck & Co., Inc., Rahway, NJ, USA

ABSTRACT

In oncology, Line of Therapy refers to the specific phase or sequence of treatment that a patient undergoes in the management of their cancer. Cancer treatment is often organized into sequential lines of therapy, each representing a distinct phase or set of interventions. However, most healthcare databases lack explicit information on treatment line of therapy. This paper introduces an innovative SAS® macro designed to depict patient treatment regimens in oncology using a defined algorithm. The algorithm initially defines the treatment regimen within a specified timeframe of the index date. Stopping drugs from the combination regimen does not advance the treatment line, but adding a new drug will start the next line of therapy. If the duration between two cycles, lacking any chemotherapy or biologic regimen, exceeds the allowable gap days, a new line of therapy is instituted. The proposed SAS® macro integrates an embedded macro to create types and flags, distinguishing various scenarios. These indicators are then utilized to subset fully defined and non-completed defined data. A loop is employed to process the remaining data, ultimately combining each defined line to capture its entire therapeutic pathway. This macro provides a comprehensive tool for analyzing real-world oncology data. The paper showcases the macro's methodology, applications, and advantages, emphasizing its potential to refine treatment regimens and improve our understanding of patient journeys in cancer care.

INTRODUCTION

In oncology, Line of Therapy (LOT) refers to the specific phase or sequence of treatment that a patient undergoes in the management of their cancer. Cancer treatment is often organized into sequential lines of therapy, each representing a distinct phase or set of interventions. However, most healthcare databases lack explicit information on treatment line of therapy. This paper introduces an innovative SAS® macro designed to depict patient treatment regimens in oncology using a defined algorithm.

This paper focuses on SAS® macro programming to define the line of therapy. The line of therapy algorithm described herein presents the general rules for defining the line of therapy in real-world oncology data. The macro, designed with multiple macro variables, offers significant flexibility to accommodate various applications of line of therapy algorithms with different parameters, including different allowable gap days [1-3]. However, discussing or modifying the line of therapy algorithm itself falls outside the scope of this paper.

LINE OF THERAPY ALGORITHM

1. A new line of therapy (LOT) is created when a continuous gap of more than &gap days (for example, gap=45 days) is observed.
    (1) To prepare the data, drug start date and drug end date should be created.
    (2) For oral agents, days of supply should be considered when calculating the drug end date.
    (3) If a drug only has the prescribed date but lacks an end date, for example, intravenous drugs, the drug_start_date and drug_end_date could be set as the same as the prescribed date.
2. A combination regimen is determined when one or more additional systemic therapy drugs are administered within &firstwindowdays (for example, 28 days) of the LOT index date.
3. A LOT will include any change of therapy as long as the new therapy is administered within &firstwindowdays (for example, 28 days) of the LOT index date
4. Dropping any drugs from the combination regimen does not advance the line of metastatic treatment. That is, continuation of a single drug from a combination regimen will not be considered as a new line of therapy.

5. If the time window without any chemotherapy/biologic regimen between two cycles exceeds the allowable gap days (for example, allowable gap days=45 days), a new line of therapy/regimen will be created.

DATA SOURCE

This macro can be applied to any database without limitations. However, certain data preparation steps must be completed before inputting data into the macro. While the explanation that follows will utilize a metastatic breast cancer population to illustrate the macro, the input data for this macro can encompass any type of oncology data. The minimum requirement for the input data is patient-level longitudinal data containing each patient's oncology treatment history. To demonstrate the functionality of the macro in this paper, a sample data set comprising metastatic breast cancer patients, along with their oncology treatments (intravenous or oral), treatment start dates, and end dates, will be utilized. The data set is sorted by ID, medication start date, and end date.

Sample data

	patient_ID	drug_date	drug	days_of_supply	class	drug_start_date	drug_end_date
1	PATID_001	05DEC2013	PACLITAXEL	.	CHEMO	12/05/2013	12/05/2013
2	PATID_001	05DEC2013	CARBOPLATIN	.	CHEMO	12/05/2013	12/05/2013
3	PATID_001	31DEC2013	CARBOPLATIN	.	CHEMO	12/31/2013	12/31/2013
4	PATID_001	31DEC2013	PACLITAXEL	.	CHEMO	12/31/2013	12/31/2013
5	PATID_001	30JAN2014	CARBOPLATIN	.	CHEMO	01/30/2014	01/30/2014
6	PATID_001	30JAN2014	PACLITAXEL	.	CHEMO	01/30/2014	01/30/2014
7	PATID_001	14SEP2015	CARBOPLATIN	.	CHEMO	09/14/2015	09/14/2015
8	PATID_001	14SEP2015	PACLITAXEL	.	CHEMO	09/14/2015	09/14/2015
9	PATID_001	08OCT2015	CARBOPLATIN	.	CHEMO	10/08/2015	10/08/2015
10	PATID_001	08OCT2015	PACLITAXEL	.	CHEMO	10/08/2015	10/08/2015
11	PATID_001	29OCT2015	CARBOPLATIN	.	CHEMO	10/29/2015	10/29/2015
12	PATID_001	29OCT2015	PACLITAXEL	.	CHEMO	10/29/2015	10/29/2015
13	PATID_001	19NOV2015	PACLITAXEL	.	CHEMO	11/19/2015	11/19/2015
14	PATID_001	19NOV2015	CARBOPLATIN	.	CHEMO	11/19/2015	11/19/2015
15	PATID_001	18APR2018	LETROZOLE	30	HORMO	04/18/2018	05/18/2018

Note: only oral drugs have non missing days of supply. The days of supply for intravenous drugs are shown as missing value.

SAS MACRO

Sample Call

```
%line0of0therapy(
        /* patient ID */
        patid=patient_ID,
        /* input data set name */
        lotindata=pat_therapy,
        /* estimated maximum number of lines, suggested to use a number slightly larger than expected */
        max_line=10,
```

```
/* variable name of drug class category, in this sample call, the variable name is set as “class” */
class=class,
/* the number of days from the first date, this window is used to determine the initial treatment regimen */
firstwindowdays=28,
/* the number of days without any regimen between two cycles */
gap=45,
/* output data set name */
lotoutdata=pat_lot
)
```

The first step in macro %line0of0therapy sorts the data, keeps necessary variables for efficient code execution, and designates the output data set as “_others”.

```
proc sort data=&lotindata.(keep=&patid. drug_start_date drug_end_date class drug) out=_others;
    by &patid. drug_start_date drug_end_date class drug;
run;
```

The data set "_others" will serve as the input data set for the following nested macro %build0lot0flag. This aspect represents one of the noteworthy features of this macro.

The macro variable "max_line" represents the maximum estimated number of lines of therapy. The macro executes a loop to identify Line 1, Line 2, Line 3, and so on. It is recommended to set a slightly larger number in order to accommodate extensive treatment history.

```
%do i=1 %to &max_line.; /*this is an estimation of the max line, it could be different per situation*/

%build0lot0flag(indata=_others,
                patid=&patid.,
                drug=drug,
                class=class,
                date=drug_start_date,
                firstwindowdays=&firstwindowdays.,
                gap=&gap.,
                outdata=_pat2);
```

The nested macro %build0lot0flag will be explained in the later section.

After executing the nested macro %build0lot0flag, the resulting data set "_pat2" will be split into two data sets. If the line number has been determined (type='0' and flag='0' in the code, the variables type and flag will be explained in the later nested macro %build0lot0flag), the corresponding data will be partitioned into data set _L&i._allrows; otherwise, the remaining data will be partitioned into data set "_others". At this point, the data set "_others" has been overwritten and will be utilized in the subsequent loop.

```
data _L&i._allrows _others(keep=&patid. drug_start_date drug_end_date class drug);
    set _pat2;
    if type='0' and flag='0' then output _L&i._allrows;
    else output _others;
run;
```

In each data set of defined line, the last record per patient is selected. The line number, line start date, and drug regimen information are then further modified.

```
data &class._Line&i.;
    set _L&i._allrows;
    by &patid.;
    if last.&patid.;
    length line $3;
    line="L&i.";
    keep &patid. firstdate drug_regimen  class_regimen line;
    rename firstdate=Line_start_date drug_regimen=Line_regimen class_regimen=trt_category_regimen;
run;
```

After completing all loops, combine the data sets for each line and output the final data set.

```
data &class._Line_1;
    set &class._Line: ;
run;
```

```
proc sort data=&class._Line_1  out=&lotoutdata. ;
    by &patid. line line_start_date;
run;
```

Final sample data set

patient_ID	Line_start_date	Line_regimen	trt_category_regimen	line
PATID_001	12/05/2013	CARBOPLATIN+PACLITAXEL	CHEMO	L1
PATID_001	09/14/2015	CARBOPLATIN+PACLITAXEL	CHEMO	L2
PATID_001	04/18/2018	LETROZOLE	HORMO	L3
PATID_001	05/22/2018	CARBOPLATIN+LIPOSOMAL DOXORUBICIN	CHEMO	L4
PATID_001	11/20/2018	OLAPARIB	OLAPA	L5
PATID_001	06/05/2019	PACLITAXEL	CHEMO	L6

Nested macro %build0lot0flag

The first data step in this nested macro involves defining the first start date of the line of therapy and the initial regimen within the first regimen window. A RETAIN statement is employed to carry over the previous row's first drug date, drug regimen, and regimen class to the next row, to determine the initial treatment regimen.

```
%**define the LOT start date and regimen within the first 28days;
%**A combination regimen is determined when one or more additional systemic therapy drugs are
administered within 28 days (4 weeks) of the first drug date;
data pat1;
    set &indata.;
    by &patid. &date.;

    retain firstdate drug_regimen class_regimen;
    format firstdate mmddyy10.  drug_regimen class_regimen $100.;

    if first.&patid. then do;
        firstdate=&date.;
        drug_regimen=strip(&drug.);
        class_regimen=strip(&class.);

    end;
    else do;
        if &date.-firstdate<=&firstwindowdays. and index(strip(drug_regimen), strip(&drug.))=0 then do;
            drug_regimen=strip(drug_regimen)||"+"||strip(&drug.);
            if index(strip(class_regimen), strip(&class.))=0 then class_regimen=strip(class_regimen)||"+"||strip(&class.);
            end;
    end;

run;
```

The second data step initially defines two variables: type1 and type2. Type1 indicates whether the gap between two consecutive records exceeds the allowable gap days, while type2 indicates whether a new medication is added to the regimen. If either criterion in type1 or type2 is met (i.e., ='1'), the flag variable and type variable are both set to 1. Since the flag and type variables are both defined in the RETAIN statement, once they are set to 1, their values persist as 1 for the remainder of the data rows. This aspect is crucial as it enables us to segregate the data into two data sets in subsequent steps: one data set containing records for which the line has been determined, and the other data set containing the remaining data for which the line has not yet been determined and will serve as input data for the next loop.

```
data &outdata.;
    set pat1;
    by &patid. &date.;

    format pre_end_date mmddyy10. type flag $1.;
    retain type flag;

    pre_end_date=lag(drug_end_date);
    if first.&patid. then do;

        pre_end_date=.;
        type='0';
        flag='0';
    end;

    gap=&date.-pre_end_date;
    if gap>&gap. then type1='1';
    else type1='0';

    if index(strip(drug_regimen), strip(&drug.))=0 then type2='1';
    else type2='0';
    if (type1='1' or type2='1') and flag='0' then do;
        type='1';
        flag='1';
    end;
run;
```

Below it shows that the data sets have been split into two parts based on the flag and type variables. The first data displayed below where type='0' and flag='0' will include the Line 1 records. The regimen of Line 1 is Carboplatin+Paclitaxel, and the line start date is 12/05/2013. The second data displayed below are the remaining records and will serve as the input data for the next loop.

patient_ID	drug	drug_start_date	drug_end_date	firstdate	drug_regimen	type	flag	gap	type1	type2
PATID_001	CARBOPLATIN	12/05/2013	12/05/2013	12/05/2013	CARBOPLATIN	0	0	.	0	0
PATID_001	PACLITAXEL	12/05/2013	12/05/2013	12/05/2013	CARBOPLATIN+PACLITAXEL	0	0	0	0	0
PATID_001	CARBOPLATIN	12/31/2013	12/31/2013	12/05/2013	CARBOPLATIN+PACLITAXEL	0	0	26	0	0
PATID_001	PACLITAXEL	12/31/2013	12/31/2013	12/05/2013	CARBOPLATIN+PACLITAXEL	0	0	0	0	0
PATID_001	CARBOPLATIN	01/30/2014	01/30/2014	12/05/2013	CARBOPLATIN+PACLITAXEL	0	0	30	0	0
PATID_001	PACLITAXEL	01/30/2014	01/30/2014	12/05/2013	CARBOPLATIN+PACLITAXEL	0	0	0	0	0

patient_ID	drug	drug_start_date	drug_end_date	firstdate	drug_regimen	type	flag	gap	type1	type2
PATID_001	CARBOPLATIN	09/14/2015	09/14/2015	12/05/2013	CARBOPLATIN+PACLITAXEL	1	1	592	1	0
PATID_001	PACLITAXEL	09/14/2015	09/14/2015	12/05/2013	CARBOPLATIN+PACLITAXEL	1	1	0	0	0
PATID_001	CARBOPLATIN	10/08/2015	10/08/2015	12/05/2013	CARBOPLATIN+PACLITAXEL	1	1	24	0	0
PATID_001	PACLITAXEL	10/08/2015	10/08/2015	12/05/2013	CARBOPLATIN+PACLITAXEL	1	1	0	0	0
PATID_001	CARBOPLATIN	10/29/2015	10/29/2015	12/05/2013	CARBOPLATIN+PACLITAXEL	1	1	21	0	0
PATID_001	PACLITAXEL	10/29/2015	10/29/2015	12/05/2013	CARBOPLATIN+PACLITAXEL	1	1	0	0	0
PATID_001	CARBOPLATIN	11/19/2015	11/19/2015	12/05/2013	CARBOPLATIN+PACLITAXEL	1	1	21	0	0
PATID_001	PACLITAXEL	11/19/2015	11/19/2015	12/05/2013	CARBOPLATIN+PACLITAXEL	1	1	0	0	0

LIMITATION

While this macro is used to build line of therapy for cancers originating from different sites, it has been observed that there are certain scenarios that are not covered by this macro. One example is when a new drug is administered in combination with an existing drug on the same date, in which case this combination should be regarded as a new line of therapy. Therefore, it is worth considering an enhancement to the existing macro.

CONCLUSION

In conclusion, the development and application of the SAS® macro presented in this paper offers a significant advancement in defining line of therapy (LOT) in the analysis of real-world oncology data. The utilization of flag and type variables to split the data into two data sets, with one being utilized in subsequent loop steps, enhances the efficiency and effectiveness of the macro, rendering it noteworthy.

REFERENCES

[1] Optum. 2018. Determining lines of therapy (LOT) in oncology in claims databases.

<https://www.optum.com/content/dam/optum3/optum/en/resources/whitepapers/

guidelines-for-determining-lines-of-therapy-whitepaper.pdf>.

[2] A.P. Abernethy, A. Arunachalam, T. Burke, C. McKay, X. Cao, R. Sorg, D.P. Carbone,

Real-world first-line treatment and overall survival in non-small cell lung cancer

without known EGFR mutations or ALK rearrangements in US community oncology

setting, PLoS ONE 12 (6) (2017) e0178420, https://doi.org/10.1371/journal.pone.0178420

[3] W Meng, W Ou, S Chandwani, X Chen, W Black, Z Cai, Temporal phenotyping by mining healthcare data to derive lines of therapy for cancer, Journal of Biomedical Informatics, (2019), www.elsevier.com/locate/yjbin

ACKNOWLEDGMENTS

Thanks to my management, Mary Anne Rutkowski (Merck & Co., Inc.) and Julien Fraisse (Aixial Group, France) for valuable comments, advice, and validation on the paper.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Yu Feng
Merck & Co., Inc.
351 North Sumneytown Pike
North Wales, PA 19454
Email: yu.feng1@merck.com