

## Identification Rule for Missing2TAs

Xi Cheng, Beigene.com

### ABSTRACT

As one of censor reasons for PFS, 2 or more consecutive missing TAs (Tumor assessments) immediately before PFS events, hereinafter referred to as Missing2TAs, is widely used. However, except relying on data collection completely, its identification rule through programing is also different from one study to another. This paper will introduce 2 comment ways that identify records with Missing2TAs and provide 2 simple examples to discuss the lost and gain of both ways.

### INTRODUCTION

The purpose of this paper is to facilitate statistical programmers understanding the definition of Missing2TAs and to introduce 2 comment ways to identify records with Missing2TAs. It will be grateful if a CRF has collected this information for a study or statistical programmers do not need to derive this censor reason for PFS. Because it is hard to take every condition into consideration and too tricky enough to identify it properly, if we totally depend on computational thinking and remote from reality - clinical trial.

To understand this, this paper will explain the definition of Missing2TAs, firstly. Although this definition may be slightly different from one company to another, we would like to use a simple rule as an example to make clear what kind of event we are dealing with. Then, two comment methods will be raised under specific scenario, 'Based on Day' and 'Based on Visit'. At this part, we try to think about programming algorithm with two different directions. After that, two examples will be displayed for discussing the lost and gain by using the two ways. Finally, we would like to provide some suggestions to statistical programmers who are weighting the advantages and disadvantages of both methods.

### DEFINITION

Missing2TAs: If patients were dead or PD (Progression Disease) immediately after missing 2 or more consecutive tumor assessments, for PFS, it is a censor but not an event.

Before applying this definition, some preconditions are supposed to be met:

1. Patient existed both evaluable baseline and at least one evaluable post-baseline tumor assessment.
2. Patient met PFS events (progression disease or death) finally.
3. Patient existed evaluable tumor assessments and met PFS events (PD or Death) before new anti-cancer therapy.

Please notice the wordings 'immediately' and 'consecutive' are most important parts for this definition. Any missing assessments happened most earlier before PFS events or only missed once tumor assessment won't be identified as Missing2TAs.

For example, a patient missed 2 tumor assessments at Visit 2 and Visit 3 each, and then, did one evaluable assessment at Visit 4, met PFS events (PD or Death) at Visit 6. In this situation, this patient did not meet this censor criteria (Missing2TAs). Furthermore, this patient may be defined as PFS event if other preconditions had been met.

### KEY POINTS FOR IDENTIFYING MISSING2TAS

1. Scheduled tumor assessments timeline: For verifying assessments, the first point we may need to know is the scheduled tumor assessments timeline for this study, to define one evaluable tumor assessment, the target visits dates, and the frequency or duration for each assessment.

2. The reference dates for identifying Missing2TAs: Since, finally, programmers need to code the identification rules, the key reference dates may be used as the key coding parameters showing in our programs. Drawn from experience, the following dates we may use:
  - a) PFS event date: First PD/Death.
  - b) The last assessment immediately before PFS event.
  - c) Reference date: Randomized date based on company's alignment.
3. Visit windows or buffer days: we did not expect all patients can complete their visits on schedule. Any unexpected issues, like long distance from clinical center, or private affairs or especially, recent years, epidemic disease (Covid-19), can stop our patients to do tumor assessments on schedule. Therefore, clinical trial studies usually provide some buffer times or days to reduce the impact of not occurring on the exact scheduled visit dates. When calculating visit dates, it will be more comprehensive to add visit windows or buffer days into each visit duration.

## **TWO METHODS FOR IDENTIFYING MISSING2TAS**

There are 2 comment ways to identify Missing2TAs. One is 'Based on Day'; another is 'Based on Visit'. For the sake of interpretation, a dummy tumor assessment timeline will be provided as the below and bother ways (Based on Day and Based on Visit) will be discussed around this timeline.

- Dummy tumor assessment timeline:

Radiologic assessment of tumor-response status per RECIST v1.1 will be performed approximately every 9 weeks ( $\pm 7$  days) for the first 54 weeks, then every 12 weeks ( $\pm 7$  days) after 54 weeks. Tumor response will be assessed by investigators.

Under this visit timeline, the result of an evaluable assessment is supposed not to be NA, NE or missing, per RECIST v1.1. The information that every 9 weeks for the first 54 weeks and every 12 weeks after 54 weeks is the visit frequency and duration,  $\pm 7$  days is the visit window for each visit.

### **BASED ON DAY**

Basically, if the duration from last assessment date before PFS events to PFS events is longer than the duration of 2 tumor assessments, then the Missing2TAs is identified. And this way (Based on Day) directly uses days to do this identification.

Set reference duration as:

- LSTTADY = (Last Assessment Date before PFS events - Randomized Date) +1.
- EVENDY = (PFS events Date - Randomized Date) +1.
- DURDY = (PFS events Date - Last Assessment Date before PFS events) +1.

Based on the dummy timeline, we need to consider the following situations:

1. PFS events occurred before week 54.
2. PFS events occurred exactly at week 66, the first VISIT after week 54.
3. The first missing assessment before week 54, but the second after week 54.
4. All missing assessments and PFS events occurred after week 54.

Furthermore, we also need to consider the visit windows for each visit. For example, when considering PFS events occurred before week 54, we need to extend the duration of VISIT to week 55 based on dummy timeline ( $\pm 7$  days). While, when consider each assessment, we also need to trace one week back and add one more week around target VISIT based on dummy timeline ( $\pm 7$  days). Then the identification rules for each situation as the following:

1. If LSTTADY  $\leq$  55 Weeks and EVENDY  $\leq$  55 Weeks, then when DURDY  $\geq$  9 Weeks + 9 Weeks + xx Weeks, then mark as Missing2TAs.

2. If 35 Weeks <= LSTTADY <= 37 Weeks, then when DURDY >= 9 Weeks + 9 Weeks + xx Weeks, then mark as Missing2TAs.
3. If 44 Weeks <= LSTTADY <= 46 Weeks, then when DURDY >= 9 Weeks + 12 Weeks + xx Weeks, then mark as Missing2TAs.
4. If LSTTADY > 55 Weeks and EVENDY > 55 Weeks, then when DURDY >= 12 Weeks +12 Weeks + xx Weeks, then mark as Missing2TAs.

Please notice, 'xx Weeks' is the buffer days for duration of 2 tumor assessments. It is usually as same as visit windows (7 days/visit window in this case). However, assigning one visit window (7 days in this case) or two visit windows (14 days in this case) or more depends on teams' decision.

The following is a sample code for reference, we treat 'xx Weeks' as two visit windows (14 days in this case):

```

data f_inmis2ta;
length inmis2ta $1.;
set inmis2ta2;
by subject;
evendy =cut-anchdt+1;
lsttady =rsdt-anchdt+1;
durdy =cut-rsdt+1;
if lsttady <= 55*7 and evendy <= 55*7 and durdy >= (9+9+2)*7 then inmis2ta="Y";
else if 35*7 <= lsttady <= 37*7 AND durdy >= (9+9+2)*7 then inmis2ta="Y";
else if 44*7 <= lsttady <= 46*7 and durdy >= (9+12+2)*7 then inmis2ta="Y";
else if lsttady > 55*7 and evendy > 55*7 and durdy >= (12+12+2)*7 then inmis2ta="Y";
format cut e8601da.;
keep subject inmis2ta;
run;

```

## BASED ON VISIT

The basic concept for this method is to deduce the day of the second tumor assessment, after the last tumor assessment before PFS event. Since, usually, tumor assessments are tested according to scheduled visits, this task equals to find out the day of 2<sup>nd</sup> scheduled visit after the last visit before PFS event. If the time to PFS events is later than this day, then the Missing2TAs is identified. Ignoring the whole duration before missing 2 tumor assessments, this method only focusses on the relationship between the last tumor assessment before PFS event and PFS event date.

The first step of this method is to create a table to list scheduled day of each target visit and 2<sup>nd</sup> visit day after the scheduled day. Based on the dummy timeline, an example (Table 1. Scheduled Visit days) as the following, please notice, the days at this step won't add visit window ( $\pm 7$  days in this case):

Visit	Scheduled Day	2 <sup>nd</sup> Visit Day after Scheduled Date
Week 0	1	127
Week 9	64	190
Week 18	127	253
Week 27	190	316

Visit	Scheduled Day	2 <sup>nd</sup> Visit Day after Scheduled Date
Week 36	253	379
Week 45	316	463
Week 54	379	547
Week 66	463	631
Week 78	547	715
Week 90	631	799

**Table 1. Scheduled Visit days**

The second step is to map unscheduled visits into scheduled visits, so that we can derive the 2<sup>nd</sup> scheduled visit after this unscheduled visit. For example (Table 2. Unscheduled Visits):

	Map to Visit	Upper	VISITDY	Lower	Scheduled Day
UNSCHEDULED VISIT	Week 36	253	284	316	253

**Table 2. Unscheduled Visits**

Above all, VISITDY = (Date of Visit - Randomized Date) +1. We need to use some techniques to find the upper and lower boundaries for each VISITDY. Upper is the scheduled day of Week 36 in this case. Lower is the scheduled day of Week 45 in this case. Since the VISITDY is closer to the scheduled day of Week 36, we map this unscheduled visit as Week 36. Then we can conject the second scheduled visit after Week 36 is Week 54 and its corresponding scheduled day is 379.

Finally, comparing the day of the 2<sup>nd</sup> visit after scheduled date, with PFS event day. If the PFS event day is beyond the day of the 2<sup>nd</sup> visit after last visit before PFS event, then missing2TAs is identified. For example, if the last tumor assessment before PFS event is the unscheduled visit mentioned at the second step, then Table 3. Target day V.S. PFS event Day:

	Map to Visit	2 <sup>nd</sup> Visit Day after Scheduled Date	EVENDY	Missing2TAs
UNSCHEDULED VISIT	Week 36	379	583	Y

**Table 3. Target day V.S. PFS event Day**

EVENDY = (PFS events Date - Randomized Date) +1. Since EVENDY > 2<sup>nd</sup> Visit Day after Scheduled Date +7 days, this patient was regarded as Missing2TAs. Please notice when comparing days, the algorithm included the visit buffer days: 7 days in this case.

The following is a sample codes for method of 'Based on Visit':

```

/**1. Scheduled Visit days**/;

data dumvis1;
length mis2cyc $50.;
do i=0 to 54 by 9;
  mis2cyc="Cycle "||strip(put(i,best.));
  mis2d=i*7+1;
  output;
end;
do i=54 to 10024 by 12;
  mis2cyc="Cycle "||strip(put(i,best.));
  mis2d=i*7+1;
  output;
end;
run;

proc sort data=dumvis1 nodupkey;
by i mis2cyc mis2d;
run;

data dumvis2;
set dumvis1;
days=lag2(mis2d);
cycles=lag2(mis2cyc);
if days^=.;
keep cycles days mis2d mis2cyc;
run;

data inmis2ta1;
merge fin_demo(in=a keep=subject dthdt anchdt) f_nactsdt f_infnpddt date_rsrc ;
by subject;
if a;
if nmiss(dthdt,infnpddt,nactsdt)<3 then do;
  cut=min(dthdt,infnpddt,nactsdt);
  if .< rsdt <min(dthdt,infnpddt,nactsdt) then output;
end;
run;

```

```

proc sort data=inmis2ta1;
  by subject rsdt;
run;

data inmis2ta2;
  set inmis2ta1;
  by subject rsdt;
  if last.subject;
  a=rsdt-anchdt+1;
run;

/**2. Map Unscheduled Visits***/;
proc sql noprint;
  create table left as
  select distinct a.subject,a.rsdt, a.a, a.InstanceName,max(b.days) as lef from inmis2ta2 as a
  left join dumvis2 as b on b.days < a.a group by subject;

  create table right as
  select distinct a.subject,a.rsdt, a.a, a.InstanceName,min(b.days) as rig from inmis2ta2 as a
  left join dumvis2 as b on b.days > a.a group by subject;

  create table fix as
  select a.subject,a.rsdt,a.a,a.InstanceName,max(b.days)as fix from inmis2ta2 as a
  left join dumvis2 as b on b.days=a.a group by subject;
quit;

data bound;
  merge left right fix;
  by subject;
  if fix=. then do;
    if abs(a-lef)>abs(a-rig) then days=rig;
    else days=lef;
  end;
  else days=fix;
  keep subject days instancename;
run;

```

```

/**3: Compare to PD/DEATH Day**;
proc sort data= bound;
  by days;
run;

proc sort data=dumvis2;
  by days;
run;

data mis2_1;
  merge bound(in=a) dumvis2;
  by days;
  if a;
run;

proc sort data= mis2_1;
  by subject;
run;

data f_inmis2ta;
  length inmis2ta $1.:
  merge mis2_1(in=a) fin_demo(in=b keep=subject dthdt anchdt) f_nactsdt f_infpddt;
  by subject;
  if a=b;

  cut=min(dthdt,infpddt)-anchdt+1;
  if cut > mis2d+7 then inmis2ta="Y";
  keep subject inmis2ta;
run;

```

## LOST AND GAIN

Although most occasions have been included into both algorithms, programs have not the ability to replace the judgement of human beings. 2 examples will be provided to display identification results when using different methods (Based on Day or Based on Visit).

### CASE 1

Following the dummy timeline above: if one patient did the last tumor assessment at 2019-08-27 and was dead at 2022-01-12 and follow the visit as Table 4. Case 1:

VISIT	Scheduled Date	Scheduled Date + 7 days	Actual Date
Week 9	2019-06-23	2019-06-30	2019-06-21
Week 18	2019-08-25	2019-09-01	2019-08-27
Week 27	2019-10-27	2019-11-03	
Week 36	2019-12-29	2020-01-05	
Death Date			2020-01-12

**Table 4. Case 1**

Based on days, the duration from 2019-08-27 to 2020-01-12 is shorter than the duration of 2 tumor assessments before Week 54 (140 days: we add 2 weeks buffer for 2 tumor assessments). Therefore, this patient won't be identified as Missing2TAs.

On the other hand, we list the scheduled date and scheduled date +7 days based on its randomized date. Using 'Based on Visit', we speculate that this patient has missed 2 scheduled tumor assessments, since the last scheduled assessment before his/her death was supposed to be 2022-01-05.

## CASE 2

Following the dummy timeline above: if one patient did the last tumor assessment at 2018-06-25 and died at 2018-11-18 and follow the visit as Table 5. Case 2:

VISIT	Scheduled Date	Scheduled Date + 7 days	Actual Date
Week 9	2018-06-29	2018-07-06	2018-06-25
Week 18	2018-08-31	2018-09-07	
Death Date			2018-11-08
Week 27	2018-11-02	2018-11-09	

**Table 5. Case 2**

This time, if we only add one week buffer for 2 tumor assessments, the duration of 2 tumor assessments is 133 days. Based on days, the duration from week 9 to death is longer than 133 days. Then, we may flag this patient as Missing2TAs.

However, based on visit timeline, this patient may just miss one tumor assessment. Then, he/she won't be identified by using 'Based on Visit' rule.

## CONCLUSION

Obviously, the method - Based on Day is more straight forward on coding, we do not need to map unscheduled visits and calculate target visit days. It is more efficient, as long as buffer days are decided, without large discrepancy between the calculated results and real world. While, according to actual clinical operation, when a subject missed tumor assessments and then met PFS event immediately, we only traced their scheduled timeline up to the last assessment before PFS event, regardless of their whole scheduled assessments. On this point, the method - Based on Visit seems to be more reasonable and useful.

## **CONTACT INFORMATION**

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

Xi Cheng  
Beigene.com  
Chengxi0404@163.com