DYNAMIC DATA MATCHING IN CLINICAL TRIAL RESEARCH

Zaizai Lu, Kendle International, Inc.
David Shen, CTG Consulting

ABSTRACT

This paper introduces a macro we created. The purpose of the macro is to find dose-related safety parameters by dynamically matching safety parameters with dose information effectively. The macro enables the user to find adverse events (or other safety parameters) within a certain period of time following each dose. It can also find all adverse events after each dose, if no time frame is specified. The macro is easy to understand and use. The user only needs to supply with a variable for safety parameter time, a dose time variable, and a time window between the safety parameter and the previous dosing, and some other simple variables. Everything else is automatically resolved through the macro.

INTRODUCTION

In clinical trial projects, we often need to combine dose administration dataset with safety datasets, such as vital signs, adverse events, ECG, and lab results. One of the common tasks is to find dose-related safety parameters. For example, a patient is given several doses at certain time points. After each dose, his vital signs are taken, or some adverse events may occur. We need to know which adverse event or vital sign parameter is associated with which dose. In the perfect world, we would hope an adverse event to occur within 3 hours of a dose, or expect vital signs to be measured 2 hours after a dose. But in the real world, most data are irregular. Some data maybe missing, some may contain duplicates. We might have some control over vital sign measurement times, but we do not have any control over adverse event occurrences. There is high unpredictability as to when an adverse event may occur. This unpredictability makes it difficult to find dose-related safety parameters. This paper presents a technique that can dynamically match safety parameters with dose information effectively. One macro program is introduced here. This macro enables the user to find all adverse events (or other safety parameters) within a certain period of time following each dose. If no time frame is defined, then all adverse events after each dose will be presented. This macro is easy to understand and use, and has great applicability in real-life clinical trial projects, especially Phase-I trials.

DATA TYPE

A variety of types of data may be applied with this method. The major types are:
1. Adverse Events
2. Vital Signs
3. ECG
4. Lab Results

DATA PROCESSING

Suppose there is a dose dataset and an adverse event dataset.

Dose Data

<table>
<thead>
<tr>
<th>OBS</th>
<th>PATNO</th>
<th>DOSEDTTM</th>
<th>DOSEAMT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0001</td>
<td>30JAN1999:08:00:01</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>0001</td>
<td>06FEB1999:08:00:00</td>
<td>30</td>
</tr>
<tr>
<td>3</td>
<td>0001</td>
<td>20FEB1999:08:00:00</td>
<td>90</td>
</tr>
<tr>
<td>4</td>
<td>0001</td>
<td>27FEB1999:08:00:02</td>
<td>120</td>
</tr>
<tr>
<td>5</td>
<td>0002</td>
<td>30JAN1999:08:01:02</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>0002</td>
<td>06FEB1999:08:01:01</td>
<td>30</td>
</tr>
<tr>
<td>7</td>
<td>0002</td>
<td>20FEB1999:08:01:00</td>
<td>90</td>
</tr>
<tr>
<td>8</td>
<td>0002</td>
<td>06MAR1999:08:01:00</td>
<td>120</td>
</tr>
<tr>
<td>9</td>
<td>0003</td>
<td>30JAN1999:08:02:01</td>
<td>10</td>
</tr>
<tr>
<td>10</td>
<td>0003</td>
<td>06FEB1999:08:02:01</td>
<td>30</td>
</tr>
<tr>
<td>11</td>
<td>0003</td>
<td>13FEB1999:08:02:01</td>
<td>60</td>
</tr>
<tr>
<td>12</td>
<td>0003</td>
<td>27FEB1999:08:02:08</td>
<td>90</td>
</tr>
<tr>
<td>13</td>
<td>0003</td>
<td>06MAR1999:08:02:01</td>
<td>120</td>
</tr>
</tbody>
</table>

This paper is composed of 6 parts. Part 1 includes the abstract and introduction. Part 2 describes the types of data file that can be analyzed using this macro. Part 3 presents some data files used in the paper to generate the sample output of the SAS® macro program. Part 4 explains the macro program in more details. Part 5 presents the source code of the macro. Part 6 exhibits some examples of the applications of the macro program and their output.
Adverse Event Data

<table>
<thead>
<tr>
<th>OBS</th>
<th>PATNO</th>
<th>AE</th>
<th>AEDTTM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0001</td>
<td>NERVOUSNESS</td>
<td>30JAN1999:06:00:00</td>
</tr>
<tr>
<td>2</td>
<td>0001</td>
<td>TACHYCARDIA</td>
<td>30JAN1999:12:15:00</td>
</tr>
<tr>
<td>3</td>
<td>0001</td>
<td>NAUSEA</td>
<td>06FEB1999:16:20:00</td>
</tr>
<tr>
<td>4</td>
<td>0001</td>
<td>DIZZINESS</td>
<td>20FEB1999:09:20:00</td>
</tr>
<tr>
<td>5</td>
<td>0002</td>
<td>HEADACHE</td>
<td>06FEB1999:14:10:00</td>
</tr>
<tr>
<td>6</td>
<td>0002</td>
<td>NAUSEA</td>
<td>06FEB1999:17:40:00</td>
</tr>
<tr>
<td>7</td>
<td>0002</td>
<td>VASODILATATION</td>
<td>20FEB1999:14:30:00</td>
</tr>
<tr>
<td>8</td>
<td>0002</td>
<td>ASTHENA</td>
<td>20FEB1999:18:20:00</td>
</tr>
<tr>
<td>9</td>
<td>0002</td>
<td>ANXIETY</td>
<td>27FEB1999:20:01:00</td>
</tr>
<tr>
<td>10</td>
<td>0002</td>
<td>CHILLS</td>
<td>28FEB1999:06:00:00</td>
</tr>
<tr>
<td>11</td>
<td>0002</td>
<td>CONSTIPATION</td>
<td>28FEB1999:22:00:00</td>
</tr>
<tr>
<td>12</td>
<td>0003</td>
<td>PALLOR</td>
<td>30JAN1999:10:17:00</td>
</tr>
<tr>
<td>13</td>
<td>0003</td>
<td>SWEATING</td>
<td>30JAN1999:12:17:00</td>
</tr>
<tr>
<td>14</td>
<td>0003</td>
<td>NAUSEA</td>
<td>06FEB1999:15:17:00</td>
</tr>
<tr>
<td>15</td>
<td>0003</td>
<td>DIZZINESS</td>
<td>06FEB1999:17:17:00</td>
</tr>
<tr>
<td>16</td>
<td>0003</td>
<td>HYPESTHESIA</td>
<td>13FEB1999:09:30:00</td>
</tr>
<tr>
<td>17</td>
<td>0003</td>
<td>PAIN</td>
<td>13FEB1999:12:30:00</td>
</tr>
<tr>
<td>18</td>
<td>0003</td>
<td>CHILLS</td>
<td>27FEB1999:16:45:00</td>
</tr>
<tr>
<td>19</td>
<td>0003</td>
<td>EUPHORIA</td>
<td>06MAR1999:10:15:00</td>
</tr>
<tr>
<td>20</td>
<td>0003</td>
<td>DIZZINESS</td>
<td>06MAR1999:18:46:00</td>
</tr>
</tbody>
</table>

Now, we need to create a new dataset to list all adverse events within a certain period of time (for instance, 6 or 12 hours following each dose administration).

EXPLANATION OF THE MACRO

Our macro %match is able to do this job effectively and efficiently. What you need to do is to define the name of the dose data set (dosedata), the name of the safety parameter data set (safedata), the time window between the safety parameter and the previous dosing (hrs). Other macro variables used in this macro include the patient id variable (patno), the variable name of the dose date and time (dosetime), the variable name of the safety parameter date and time (safetime), and the output dataset name (outdata). Both dosetime and safetime are numeric. In theory, patients are supposed to take doses on schedule. However, some patients may fail to follow the schedule and miss several doses. So actually, patients may not have the same number of doses. The first step in this macro is to capture as a macro variable the largest number of doses among all the patients. The output of PROC FREQ contains one observation per PATNO, with the variable COUNT indicating the number of doses each patient receives. Because of the ORDER=FREQ option on the PROC FREQ statement, the variable COUNT is put in descending order, with the first observation having the largest value. Using a DATA _NULL_ step and selecting _N_=1, we are able to create a macro variable to get the maximum number of doses in the whole dose dataset. The next step PROC TRANSPOSE creates a temporary dataset containing one observation for each patient with all the dosing data. After one-to-many merging, the time difference can be conveniently calculated by subtracting safety parameter time from the dosing time. To be easily understood, we can deal with one dose each time through a loop from 1 to the largest dosing number. The resulting data set will be created by comparing the calculated time difference with the defined time window. Each dose will generate a subset of datasets after the process. When the looping is done, all resulting datasets with the expected safety parameter data will then be merged together to generate the final output dataset with the name you wish to call.

SOURCE CODE OF THE MACRO PROGRAM

```
%macro match (dosedata=, safedata=, hrs=, patno=, dosetime=, safetime=, outdata=);
  proc sort data=&dosedata;
    by &patno &dosetime;
  run;
  proc freq data=&dosedata order=freq noprint;
    table &patno /out=temp1;
  run;
  data _null_; 
    set temp1; 
    if _n_=1 then call symput('maxdose', count);
  run;
  proc transpose data=&dosedata out=temp2
    (drop=_name_) prefix=dosen;
    by &patno; 
    var &dosetime;
  run;
  %local num;
  %do num=1 %to &maxdose;
    proc sort data=temp2
      out=dose&num(keep=&patno dosen&num);
    by &patno;
    run;
    data ndata&num ;
      merge &safedata(in=a) dose&num(in=b);
      by &patno;
      if a and b;
      dosen=&num;
      hrs=(&safetime - dosen&num)/3600;
      &dosetime=dosen&num;
    %if &hrs ne %then %do;  *ALL DATA WILL BE COLLECTED FOLLOWING EACH DOSE IF NO HOURS IS SET; 
      if 0=<hrs<=&hrs then output ndata&num;
    %end;
  %end;
%macro end;
```
%else if &hrs eq %then %do; *ALL DATA WILL BE COLLECTED FOLLOWING EACH DOSE IF NO HOURS IS SET;

if 0<&hrs then output ndata;&
%end;
drop dosen&num;
run;
%end;

data &outdata;
set ndata1;
run;

%do num=2 %to &maxdose;
set ndata1;
run;
%end;

%if &hrs eq %then %do;
*ALL DATA
%end;

If we leave the time window missing, then the dataset will look like this, with all the adverse events following each dose:

%match(dosedata=dose, safedata=adverse, hrs=, patno=patno, dosetime=dosedttm, safetime=aedttm, outdata=c);

c) Adverse event dataset with no time difference defined

If we call this macro with different time windows, %match(dosedata=dose, safedata=adverse, hrs=6, patno=patno, dosetime=dosedttm, safetime=aedttm, outdata=c);

the following 2 datasets would be generated.

a) Adverse event dataset with time difference of 6 hours

b) Adverse event dataset with time difference of 12 hours
REFERENCES


TRADEMARKS

SAS is a registered trademark or trademark of SAS Institute Inc. in the USA and other countries.
® indicates USA registration.

CONTACT INFORMATION

For additional information, contact the author at:

Zaizai Lu
Kendle International, Inc.
1200 Carew Tower
441 Vine Street
Cincinnati, OH 45202
lu.zaizai@kendle.com