INTRODUCTION

In the clinical trial data processing, TREATMENT GROUP (variable TX) and ELIGIBILITY (variable ELIG) data are frequently used for safety and efficacy analysis. It is time consuming to merge the data sets containing TX and ELIG data with an analyzed data set. This paper will show you how can build TX and ELIG formats with control data sets to access TX and ELIG data efficiently.

CREATING AND USING CONTROL DATA SET

A control data set used in a CNTLIN= option is a SAS data set that the FORMAT procedure can use to construct formats and informats. Creating control data sets is to alter an existing SAS data set so that it conforms to control data set structure. This is demonstrated in the example that follow.

Suppose the data set TXADMIN contains TX data and the data set MKELIG contains ELIG data, they are:

<table>
<thead>
<tr>
<th>PATID</th>
<th>TX</th>
</tr>
</thead>
<tbody>
<tr>
<td>101</td>
<td>1</td>
</tr>
<tr>
<td>102</td>
<td>1</td>
</tr>
<tr>
<td>103</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PATID</th>
<th>ELIG</th>
</tr>
</thead>
<tbody>
<tr>
<td>101</td>
<td>Eligible</td>
</tr>
<tr>
<td>102</td>
<td>Ineligible</td>
</tr>
<tr>
<td>103</td>
<td>Violator</td>
</tr>
</tbody>
</table>

The following statements create control data sets TXCNT and ELIGCNT:

```sas
data TXCNT;
retain fmtname 'TX'
   type 'N';
set TXADMIN;
start=patid;
select;
   when tx=1 then label='Experimental';
   when tx=2 then label='Control';
   otherwise label=' '; 
end;
run;

data ELIGCNT;
retain fmtname 'ELIG'
   type 'N';
set MKELIG(rename=(patid=start elig=label));
run;
```

Once the control data sets create, the TX and ELIG formats can be built by using FORMAT procedure with CNTLIN option as follows:

```sas
libname library 'SAS-data-library';
proc format library=library.formats2
cntlin=TXCNT;
   ( or cntlin=ELIGCNT);
run;
```

SPECIFYING ANOTHER PERMANENT FORMAT CATALOG

If you specify LIBRARY=LIBRARY in the PROC FORMAT statement, all associated formats and informats with a study raw data normally are stored in the permanent
catalog LIBRARY.FORMATS. If you reference any permanent formats or informats in a subsequent DATA or PROC step, the SAS system automatically searches the LIBRARY FORMATS catalog the permanent formats or informats.

Considering the data confidentiality and formats library maintenance, TX and ELIG formats can be stored in another permanent format catalog LIBRARY.FORMATS2. FMTSEARCH=system option enables you to name additional catalogs that SAS system will automatically search.

AN EXAMPLE OF ACCESSING TX AND ELIG DATA:

```sas
options fmtsearch=(library.formats2);

data DEMOG;
set SSD.DEMOG;
TX=put(patid, TX.);
if put(patid,elig) in ('eligible','violator');
run;
```

Apparently, you do not need to find the data sets containing TX and ELIG data. With the created TX and ELIG formats, you can easily calculate how many patients are assigned randomly to experimental group, how many patients violate protocol criteria and how many patients are appropriate for ITT(intention-to-treat) analysis.

SUMMARY

It is obvious that building formats and informats to access necessary data is more efficient than merging process. The CNTLIN=control data set provides a way to construct formats and informats without specify all the informat in value, picture, or invalue statements.

ACKNOWLEDGMENTS

The author greatly appreciates Monte Jarvis and Denis Gungor for their continual support and constructive suggestion on this paper.

SAS is a registered trademark of SAS Institute Inc. in the USA and other countries.

Author Contact

Frank Liu

E-mail: fliu@world.std.com