Paper: 089-2012

Utilize Dummy Datasets in Clinical Statistical Programming
Amos Shu, Endo Pharmaceuticals., Chadds Ford, PA

ABSTRACT
Due to collectability or other issues, some clinical trial reporting tables like physical examination, demographic characteristics, and some efficacy tables usually need to be partially made up in some way in the real clinical practice world. Creating dummy datasets is an effective way to improve programming efficiency in these situations. This paper discusses six ways to utilize dummy datasets in clinical statistical programming.

INTRODUCTION
Typically, tables like physical examination, demographic characteristics, and some efficacy tables need to be partially made up because not all data is collectable in the real clinical practice world. Utilizing dummy datasets is an effective way to improve programming efficiency in these situations. There are six ways to utilize dummy datasets in clinical statistical programming.

All programs presented in this paper were developed in Server SAS$^{®}$ 9.2 in the Windows environment.

1. USE DUMMY DATASETS TO MAKE UP PHYISICAL EXAMINATION TABLES
Physical examination tables usually contain all medical scenarios such as 'Normal', 'Abnormal', and 'Not Done' for each organ regardless of if the actual data is available. They appear like the following table.

_____

	Placebo (N=45)	Treat1 (N=43)	Treat2 (N=43)
Heart			
Normal	8	9	11
Abnormal	0	0	0
Not Done	0	0	0
Lungs/Chest			
Normal	7	8	10
Abnormal	0	1	0
Not Done	1	0	1

......
_____

In real practice, not all subjects have all findings for all organs. For example, there may be only 'Normal' Heart data, no 'Abnormal' or 'Not Done' data available for Heart. In this case, creating a dummy dataset will be very helpful to generate this type of tables.

```
DATA dummy;
     DO x = 1 to 9;
          DO y = 1 to 3;
          Output;
          END;
     END;
RUN;
```
Then associate x with each organ and associate y with the values - 'Normal', 'Abnormal', and 'Not Done'. Finally merge with raw data, the desired table will be easily generated.

1

Body System	Findings
General Appearance	Normal
Lungs/Chest	Normal
Lungs/Chest	Abnormal
Heart	Normal
Heart	Abnormal
…	…

x	y	Body System	Findings
1	1	General Appearance	Normal
1	2	General Appearance	Abnormal
1	3	General Appearance	Not Done
2	1	Lungs/Chest	Normal
2	2	Lungs/Chest	Abnormal
2	3	Lungs/Chest	Not Done
3	1	Heart	Normal
3	2	Heart	Abnormal
3	3	Heart	Not Done
…	…	…	…

2. USE DUMMY DATASETS TO MAKE UP DEMOGRAPHIC CHARACTERISTICS TABLES

Only a few subjects are enrolled at the beginning of any clinical study, so not all races are available. Sometimes even in the end of a study, not all races are represented. We need to display all race information in the summary of a demographic characteristics table as follows:

```
                                         Placebo        Treat1         Treat2
                                         (N=45)         (N=43)         (N=43)
_____

Race
    Caucasian                              28             29             27
    African American                        0              3              1
    Asian                                  10             11              9
    American Indian or Alaska native        0              0              0
    Native Hawaiian or Pacific Islander     0              0              0
    Other                                   1              0              1


......
_____
```

To avoid possible error and improve programming efficiency, it would be better to use a dummy dataset to summarize races at the beginning of programming. With time and more subjects enrolled in the study, if new race becomes available, you do not have to modify the SAS program.

3. USE DUMMY DATASETS TO OPTIMIZE DATASETS

In efficacy analysis, statisticians often perform various subgroup analyses such as geographic subgroups – US, Europe, and Asia, sex subgroups – male and female, race subgroups – Asian, Non-Asian, smoking status subgroups- never smoker and current smoker, age subgroups – age <=65 and age >65. To perform those analyses, a macro is usually used to handle the SAS job. However, some subgroups may not have all treatment values available, e.g., subgroup A may contain treatment 1 and 2 only but not treatment 3 and subgroup B may contain treatment 1 and 3 only but not treatment 2. Without the help of a treatment dummy dataset, the macro would fail.

```
        DATA  dummytrt ;
              Treat1 = . ; treat2= .; treat3 = .;
        RUN;

        DATA  SubEff_d;
           MERGE  dummytrt   SubEff ;
        RUN;
```

Please note that the treatment dummy dataset must be placed as the first dataset in the MERGE statement to avoid overwriting the treatment values in original dataset. After the merge, every subgroup contains all treatments. The macro can then perform efficacy analyses without any problem.

4. USE DUMMY DATASETS TO STACK DATASETS IN MACRO

If you do not know exactly how many sub datasets are available to stack together for analysis, it will be impossible to use a single SET statement to stack all sub datasets in a DO loop macro. A dummy dataset is helpful here.

```
DATA  comb ;
      x = 0 ;
RUN;

%DO i= 1 %TO &sqlobs.;
      ...
      DATA  sub&i. ;
             ... ;
      RUN;

      DATA  comb ;
         SET  comb   sub&i.  ;
      RUN ;
%END;

DATA  final;
   SET  comb;
   IF x = 0  THEN DELETE ;
RUN;
```

5. USE DUMMY DATASETS TO IMPUTE LAST OBSERVATION CARRIED FORWARD (LOCF)

There are normally two types of situations for LOCF – In the example below, a dataset may contain no data for a visit number, or both the visit number and data are missing for a subject. A dummy dataset would be useful to perform a LOCF job in such situations.

SUBJECT	VISIT	AVAL
101	Visit 9	10.8
102	Visit 9	9.5
102	Visit 10	10.1
102	Visit 11	
…		
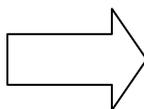
SUBJECT	VISIT	AVAL
101	Visit 9	10.8
101	Visit 10	10.8
101	Visit 11	10.8
102	Visit 9	9.5
102	Visit 10	10.1
102	Visit 11	10.1
…		

```
DATA  dummy ;
      SET  lockbase;
      DO  AVISITN = 9, 10, 11 ;
            output ;
      END;
RUN;

DATA  lockdata2 ;
   MERGE  dummy   lockdata ;
   BY  USUBJID   AVISITN ;
RUN;
```

After making up visit values, use RETAIN statements to fill out all missing data by LOCF.

6. USE DUMMY DATASETS TO ADD EXTRA LINES IN RTF OUTPUT

Dummy datasets can help improve the appearance of a table. For example, a Visit dummy dataset can help add visit lines (week 1, week2, etc.) and blank lines as in the output table below. Though compute block can add extra lines between visits, dummy dataset is an alternative way to do this.

3

visit	stat
Week 1	N
Week 1	Mean
Week 1	SD
Week 1	Median
Week 1	Minimum
Week 1	Maximum
Week 2	N
Week 2	Mean
Week 2	SD
Week 2	Median
Week 2	Minimum
Week 2	Maximum
…	
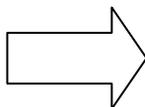
visit	stat
Week 1	Week1
Week 1	N
Week 1	Mean
Week 1	SD
Week 1	Median
Week 1	Minimum
Week 1	Maximum
Week 1	
Week 2	Week2
Week 2	N
Week 2	Mean
Week 2	SD
Week 2	Median
Week 2	Minimum
Week 2	Maximum
Week 2	
…	

```
_____

                         Placebo            Treat1            Treat2
                         (N=45)             (N=43)            (N=43)
_____

Week 1
    N                      103                103               101
    Mean                    36.2               37.4              36.6
    SD                      10.25              10.43              9.77
    Median                  36.0               37.0              36.0
    Minimum                  2                 11                 6
    Maximum                 59                 64                60

Week 2
    N                      103                102               101
    Mean                    35.9               37.1              36.3
    SD                      10.15              10.03             10.07
    Median                  36.0               37.0              36.0
    Minimum                  3                 10                 6
    Maximum                 58                 64                59

......
_____
```

CONCLUSION

Creating a dummy dataset is easy and simple. A simple dummy dataset is extremely helpful to generate physical examination, demographic characteristics, and some efficacy tables. Understanding when and how to create a dummy dataset in your programming process is important in clinical statistical programming work.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Please contact the author at:
Amos Shu
Endo Pharmaceuticals
100 Endo Blvd.
Chadds Ford, PA 19317
Email: shu.amos@endo.com

TRADEMARK INFORMATION

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.