PharmaSUG 2018 - Paper EP-23

Summary Table for Displaying Results of a Logistic Regression Analysis

Lori S. Parsons, ICON Clinical Research, Medical Affairs Statistical Analysis

ABSTRACT

When performing a logistic regression analysis (LR) for a study with the LOGISTIC procedure, analysts often want to summarize the results of the analysis in a compact table. A combination of SAS® DATA step, SAS macro, and SAS Output Delivery System code is presented here as a method to create and display a summary LR table. The summary table includes the odds ratio estimate (OR), 95% confidence interval (CI), and P value for each covariate included in the LR model. The reference group (Ref) of each covariate included in the model is indicated in the table. The total number of observations (N) for each covariate level is also included in the table. Additionally, a table footnote is generated, by macro variable code, to describe the total number of observations read and the total number of observations used in the LR model.

INTRODUCTION

The SAS Output Delivery System (ODS) allows analysts to create a SAS dataset containing the results of a SAS procedure. In the method presented here, the ODS output objects ODDSRATIOS, PARAMETERESTIMATES, and NOBS are used to select the desired results from PROC LOGISTIC. The ODS output object ONEWAYFREQS is used to select the desired results from the FREQUENCY procedure.

The dataset NOBS will be used to create a footnote containing the number of observations read and the number of observations included in the LR model.

The summary table will be printed to a Rich Text Format (RTF) file using the REPORT procedure. The ASIS=ON option will be used within PROC REPORT to specify that leading spaces will be honored allowing for analyst controlled character indentation.

METHOD

STEP 1: SAS® SETUP

Setup the SAS environment:

```
ods escapechar  = '^';
%let indat1     = SampleData_P01;
%let outdat1    = SampleTable;
%let psugpath   = U:\PharmaSUG_2018;
libname saslib  "&psugpath.\saslib";
```

STEP 2: CREATE A PATIENT LEVEL INDICATOR VARIABLE FOR LR MODEL OUTCOME AND COVARIATE DATA ALL NON-MISSING

Create the indicator variable ind_lrcompl, which will be used to determine the N for each covariate level in the summary table. In the array ARcompl(*) statement, include the model outcome variable and all of the model covariate variables:

```
data cohort01;
  set saslib.&indat1.;
```

```
  ind_lrcompl = 1;
  label ind_lrcompl = "LR model outcome and covariate data all non-
missing";

  array ARcompl(*) cabg gender age_ge65 mhchf mhdm mhhyp mhhyper mhpad
                   mhpci mhstroke cctotal;
  do i = 1 to dim(ARcompl);
     if ARcompl(i) eq . then ind_lrcompl = 0;
  end;
run;
```

STEP 3: CREATE A BASE TABLE

Create a base table with columns for covariate labels, values, and Ns:

```
data base01;
   length charcol $100 col1 $8;
   label charcol = "Covariate";
   label col1 = "N";
run;
```

STEP 4: POPULATE THE BASE TABLE

Run the macro %MPopulate (covarl=, covar=) (see Appendix) for each covariate that will be included in the LR model. This will populate the base table with covariate labels, covariate values, and cohort Ns.

Table 1 shows the base table after populating it with the first three covariates. The method for indenting the covariate value is described in Step 13 of this paper.

covarl= Specifies the covariate label for the header row in the table.

covar= Specifies the variable name of the covariate that will be included in model. The covariate values (Male, Female, etc.) were assigned by a format statement when the dataset was created.

```
%MPopulate(covarl=Gender,                                    covar=gender);
%MPopulate(covarl=Age,                                       covar=age_ge65);
%MPopulate(covarl=Total number of comorbidities,             covar=cctotal);
%MPopulate(covarl=Medical history of CHF,                    covar=mhchf);
%MPopulate(covarl=Medical history of diabetes mellitus,      covar=mhdm);
%MPopulate(covarl=Medical history of hypertension,           covar=mhhyp);
%MPopulate(covarl=Medical history of hyperlipidemia,         covar=mhhyper);
%MPopulate(covarl=Medical history of peripheral artery disease,
                                                             covar=mhpad);
%MPopulate(covarl=Medical history of percutaneous coronary intervention,
                                                             covar=mhpci);
%MPopulate(covarl=Medical history of stroke,                 covar=mhstroke);
```

Table 1. Base Table Populated with First Three Covariates

Covariate	N
Gender	
Male	259
Female	199
Age	
<65	258
>=65	200

Covariate	N
Total number of comorbidities	
1	46
2	68
3+	344

STEP 5: SETUP BASE TABLE FOR MERGE WITH LR MODEL DATA

Create variable rownum, which will be used to sort the final summary table of combined base and LR model data.

Create variable ind_refrow, which is an indicator for the covariate reference category. The reference category is specified in the PROC LOGISTIC code (see Step 6).

Create variable mergenum, which will be used to merge the LR model OR, CI, and P value data.

Table 2 shows the resulting dataset (base03) for the first three covariates:

```
data base02 (keep = charcol col1 ind_refrow rownum);
   set base01;
   where charcol ne "";

   rownum + 1;
   if rownum eq 1 then tcol1 = col1;
      else if rownum gt 1 and tcol1 = "" then ind_refrow = 1;
   tcol1 = col1;

   retain tcol1;
run;

data base03 (drop = tmergenum);
   set base02;

   if rownum eq 1 then tmergenum = 0;

   if col1 ne "" and ind_refrow = . then
      do;
         mergenum = tmergenum + 1;
         tmergenum = tmergenum + 1;
      end;
      else mergenum = .;

   retain tmergenum;
run;
```

Table 2. Base Table Setup for Merge with LR Data

Covariate	N	rownum	ind_refrow	mergenum
Gender		1	.	.
Male	259	2	1	.
Female	199	3	.	1
Age		4	.	.
<65	258	5	1	.
>=65	200	6	.	2

Covariate	N	rownum	ind_refrow	mergenum
Total number of comorbidities		7	.	.
1	46	8	1	.
2	68	9	.	3
3+	344	10	.	4

STEP 6: RUN THE LR MODEL

Use the ODS output code for objects ODDSRATIOS (OR and CI data), PARAMETERESTIMATES (P value), and NOBS (number of observations read and number of observations used):

```
ods output OddsRatios=OR01;
ods output ParameterEstimates=PE01 (where=(variable ne "Intercept"));
ods output NObs = NObs (where=(label eq "Number of Observations Used"));
proc logistic data=cohort01;
  class cabg          (ref='No')
        gender        (ref='Male')
        age_ge65      (ref='<65')
        cctotal       (ref='1')
        mhchf         (ref='No')
        mhdm          (ref='No')
        mhhyp         (ref='No')
        mhhyper       (ref='No')
        mhpad         (ref='No')
        mhpci         (ref='No')
        mhstroke      (ref='No')
        /param=ref;
  model cabg = gender age_ge65 cctotal mhchf mhdm mhhyp mhhyper mhpad mhpci
               mhstroke;
run;
ods output close;
quit;
```

STEP 7: SETUP LR DATASETS FOR MERGE WITH BASE TABLE

Setup LR model OR and CI results for merge:

```
data OR02;
   set OR01;
   mergenum + 1;
run;
```

Setup LR model P value results for merge:

```
data PE02;
   set PE01;
   mergenum + 1;
run;
```

STEP 8: MERGE LR DATASETS AND FORMAT OR, CI, AND P VALUE DATA

Merge the LR datasets and create the table columns or, ci, and pval which include the formatted model results.

The variable ClassVal0 in the LR dataset PE02 contains the covariate value labels, and corresponds to the charcol variable in the base table.

Table 3 shows the results of the merge and formatted data:

```
data LR01;
   merge OR02 PE02;
   by mergenum;
   length charcol $100 or ci $30 pval $10;


   or = put(oddsratioest, 6.2);
   ci = "(" || put(lowercl,5.2) || ", " || put(uppercl,5.2) || ")";
   pval = put(probchisq,6.4);
   charcol = ClassVal0;

   label charcol = "Covariate";
   label or = "Odds Ratio";
   label ci = "95% Confidence Interval";
   label pval = "P value";
run;
```

Table 3. Table Columns with LR Formatted Data

mergenum	Covariate	Odds Ratio	95% Confidence Interval	P value
1	Female	0.28	(0.10, 0.81)	0.0181
2	>=65	2.16	(0.90, 5.16)	0.0855
3	2	1.50	(0.13, 17.01)	0.7523
4	3+	1.64	(0.19, 14.20)	0.6597
5	Yes	1.47	(0.26, 8.37)	0.6734
6	Yes	1.45	(0.59, 3.56)	0.4221
7	Yes	1.45	(0.44, 4.71)	0.5433
8	Yes	6.63	(2.17, 20.25)	0.0003
9	Yes	1.22	(0.22, 6.72)	0.8265
10	Yes	2.68	(0.82, 8.72)	0.1002
11	Yes	1.79	(0.31, 10.26)	0.5121

STEP 9: MERGE FORMATTED LR DATASETS WITH BASE TABLE

Sort base table:

```
proc sort data = base03 out = sbase03;
   by mergenum;
run;
```

Sort LR results:

```
proc sort data = LR01 out = sLR01;
  by mergenum;
run;
```

Merge base table and LR results:

```
data table01 (keep = charcol col1 rownum ind_refrow mergenum or ci pval);
   merge sbase03 sLR01 (drop = charcol);
   by mergenum;
run;
```

STEP 10: ASSIGN LABELS (REF) TO THE COVARIATE REFERENCE CATEGORY

Where the indicator variable ind_refrow = 1 (see Step 5), assign the label "(Ref)" to the table columns or, ci, and pval:

```
data table02;
   set table01;

   if ind_refrow eq 1 then
      do;
         or = "(Ref)";
         ci = "(Ref)";
         pval = "(Ref)";
      end;
run;
```

STEP 11: SORT AND KEEP FINAL TABLE COLUMNS

Sort the final table dataset by rownum (see Step 5) and keep final table columns:

```
proc sort data = table02 out = &outdat1. (keep = charcol col1 or ci pval);
   by rownum;
run;
```

STEP 12: CREATE GLOBAL MACRO VARIABLE FOR TABLE FOOTNOTE

Use the dataset NOBS created in Step 6 to create a global macro variable (&fn.) containing the number of observations read and the number of observations included in the LR model. This global macro variable will be rendered as a footnote in the LR summary table. The code is setup to format Ns <=99999, however it can easily be modified to format larger Ns:

```
%global fn;
data _null_;
  length fn_le999 $110;
  length fn_le9999 $110;
  length fn_le99999 $110;
  set NObs;

  if NObsRead le 999 then
    do;
       fn_le999 = "Note: There were " || put(NObsRead,3.0) || "
observations read, and "
       || put(NObsUsed,3.0) || " observations used in the logistic
regression analysis.";
       call symput("fn",fn_le999);
    end;

    else if NObsRead le 9999 then
       do;
          fn_le9999 = "Note: There were " || put(NObsRead,4.0) || "
observations read, and "
```

```
          || put(NObsUsed,4.0) || " observations used in the logistic
regression analysis.";
          call symput("fn",fn_le9999);
       end;

    else if NObsRead le 99999 then
       do;
          fn_le99999 = "Note: There were " || put(NObsRead,5.0) || "
observations read, and "
          || put(NObsUsed,5.0) || " observations used in the logistic
regression analysis.";
          call symput("fn",fn_le99999);
       end;
run;
```

STEP 13: PRINT LR SUMMARY TABLE TO AN RTF FILE

The covariate values (Male, Female, etc.) are setup for indenting in the following two places:

1. In Macro %MPopulate (see Appendix), leading spaces are added to the covariate value:

```
charcol = "     " || F_&covar.;
```

2. In PROC REPORT, the ASIS=ON option specifies that leading spaces will be honored allowing for analyst controlled indentation:

```
define charcol  / display style(column)=[just=left asis=on cellwidth=3in];
```

The customized footnote containing the number of observations read and the number of observations included in the LR model (global macro variable &fn.) is rendered with the ODS RTF TEXT= code:

```
ods rtf text = "^S={outputwidth=100% just=l font_face='arial'
font_size=1}{&fn.}";
```

ODS RTF and PROC REPORT code to print the LR summary table to an RTF file:

```
ods rtf file = "&psugpath\sasout\SampleTable.rtf"
               startpage = no notoc_data;
title1 font=arial height=10 pt " Table 4.  Logistic Regression Model of
Predictors of Coronary Artery Bypass Surgery^{super a}";

proc report data = &outdat1. nowd headline missing split="@"
  style(column)={foreground=black background=white font_size=8pt
     font_face="Arial" vjust=c just=c}

  style(header)={foreground=black background=white font_size=8pt
     font_weight=light font_face="Arial" vjust=c just=c};

  column charcol col1 or ci pval;
  define charcol  / display style(column)=[just=left asis=on
     cellwidth=3in];
  define col1/ display style(column)=[just=center cellwidth=1in];
  define or--pval/ display style(column)=[just=center cellwidth=1in];
run;
ods rtf text = "^S={outputwidth=100% just=l font_face='arial'
font_size=1}{&fn.}";
ods rtf close;
```

RESULTS

Table 4. Logistic Regression Model of Predictors of Coronary Artery Bypass Surgery[a]

Covariate	N	Odds Ratio	95% Confidence Interval	P value
Gender				
Male	259	(Ref)	(Ref)	(Ref)
Female	199	0.28	(0.10, 0.81)	0.0181
Age				
<65	258	(Ref)	(Ref)	(Ref)
>=65	200	2.16	(0.90, 5.16)	0.0855
Total number of comorbidities				
1	46	(Ref)	(Ref)	(Ref)
2	68	1.50	(0.13, 17.01)	0.7523
3+	344	1.64	(0.19, 14.20)	0.6597
Medical history of CHF				
No	442	(Ref)	(Ref)	(Ref)
Yes	16	1.47	(0.26, 8.37)	0.6734
Medical history of diabetes mellitus				
No	344	(Ref)	(Ref)	(Ref)
Yes	114	1.45	(0.59, 3.56)	0.4221
Medical history of hypertension				
No	99	(Ref)	(Ref)	(Ref)
Yes	359	1.45	(0.44, 4.71)	0.5433
Medical history of hyperlipidemia				
No	251	(Ref)	(Ref)	(Ref)
Yes	207	6.63	(2.17, 20.25)	0.0003
Medical history of peripheral artery disease				
No	440	(Ref)	(Ref)	(Ref)
Yes	18	1.22	(0.22, 6.72)	0.8265
Medical history of percutaneous coronary intervention				
No	430	(Ref)	(Ref)	(Ref)
Yes	28	2.68	(0.82, 8.72)	0.1002
Medical history of stroke				
No	447	(Ref)	(Ref)	(Ref)
Yes	11	1.79	(0.31, 10.26)	0.5121

Note: There were 520 observations read, and 458 observations used in the logistic regression analysis.
a. Data presented in PharmaSUG sample table are for display purposes only and do not represent actual study data.

CONCLUSION

The method presented here allows analysts to create a summary table of logistic regression results. Included in the table are the cohort Ns for each covariate level, odds ratio, 95% confidence interval, and P value. The reference category for the model is identified and a footnote describing the number of observations read and number of observations used is provided.

CONTACT INFORMATION

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

Lori S. Parsons
ICON Clinical Research, Medical Affairs Statistical Analysis
Lori.Parsons@iconplc.com

APPENDIX

```
/* ********************************** */
/* Macro to populate base table with covariate labels and values, and Ns */
/* covarl= Specify covariate label for row in table */
/* covar= Specify covariate name that will be used in LR model */
/* ********************************** */
%Macro MPopulate(covarl=, covar=);
ods select OneWayFreqs;
ods output OneWayFreqs = &covar.01;
proc freq data = Cohort01;
   table &covar.;
   where ind_lrcompl eq 1;
run;
ods output close;
quit;

data &covar.02 (keep = charcol col1);
   set &covar.01;
   length charcol $100 col1 $8;

   row + 1;
   if row eq 1 then
      do;
         charcol = "&covarl.";
         col1 = "";
         output;
      end;

   charcol = "     " || F_&covar.;
   col1 = put(frequency,8.0);
   output;
run;

data base01;
   set base01 &covar.02;
run;
%Mend MPopulate;
```