

VALIDATION: Let SAS® do the comparisons for you.

Lara E.H. Guttadauro, Kendle International Inc., Cincinnati, OH

ABSTRACT

Have you ever been asked to validate a listing? It can be a tedious, boring and sometimes inaccurate process if comparisons are done by hand. I have developed a procedure to let SAS do those comparisons. First, the output from the listing needs to be read into a SAS data set. Next, the validation results need to be calculated (you need to do this anyway) and then turned into a SAS data set with the same layout and properties as the one created from the original output. Last, SAS compares the original versus validation data sets by using PROC COMPARE. The results are concise, quick, accurate and 100% complete.

INTRODUCTION

The assignment is to validate a listing that a co-worker has created. My experience of this process has been to take a random subset of the data in the listing and check by hand to make sure the results are correctly portrayed in the output. This labor intensive process looks for inconsistencies by visually examining the original and validation outputs. I despise this assignment! So I have created a way to utilize SAS when comparing the original and validation output.

READ IN THE OUTPUT TO BE VALIDATED

The first step to letting SAS do comparison work for you is to convert the listing output into a SAS data set. Distinct columns should define the output and each observation in the data set needs to be uniquely identified.

See Appendix 1 for an example of a listing output.

In order to capture all the information in the listing, every column of the listing is read in without skipping any spaces. Information that crosses columns will be captured with the variable LINE. Once the output has been read into SAS, the data are manipulated to keep only the rows of the output containing data to be validated. Typically rows containing information such as titles, headers and footnotes are deleted and the content of these rows must be reviewed by hand.

```
*****;
** Read in existing table output and **;
** create a data set for comparison. **;
*****;

data ORIGINAL;
  infile "F:\KENDLE\PHARMSUG\OUTPUT\L_DEMO.LIS"
  missover lrecl=120 pad;

  ** Read in every column of data in the **;
  ** listing - do not skip any spaces **;
  input COL1 $ 1-9 COL2 $ 10-19 COL3 $ 20-30
        COL4 $ 31-38 COL5 $ 39-45 COL6 $ 46-57
        COL7 $ 58-82 COL8 $ 83-92 COL9 $ 93-102
        COL10 $ 103-120 @1 LINE $CHAR120.;

  ** These are the rows of the table **;
  ** that contain data to be deleted **;
  if COL1 in (' ' 'PATIENT' 'ID'
             '-----' '(a) Age i' '* At time')
  then delete;
  else if index(LINE, '.SAS')>0 and
        index(LINE, 'Page')>0 then delete;

  ** Identify the site and remove extra info **;
  if index(LINE, '<<<')>0 and index(LINE, '>>>')>0
  then do;
    SITE=substr(LINE, 24, 3);
    delete;
  end;
  retain SITE;
  drop LINE;
run;

proc sort data=ORIGINAL;
  by SITE COL1;
run;
```

The data set ORIGINAL is simply an electronic version of the data displayed in the listing output. I have found that making all the variables character is the simplest approach.

See Appendix 2 for a proc print of the ORIGINAL data set.

CREATE THE VALIDATION DATA SET

The second step to letting SAS do comparison work for you is to create a validation data set. As the validator, you will need to independently program the same information that is in the original output and then create a data set containing this information. All the variables in your VALIDATE data set will need to be character since that is how the original output was read into SAS. You must also be careful to give the variables the same names and set the data set up identically to the one you created for the original output.

USE PROC COMPARE TO LET SAS DO THE COMPARISONS

The final step is to use PROC COMPARE to let SAS compare the ORIGINAL and VALIDATE data sets for you. This step includes the actual comparison, outputting the comparison results and printing the validation data set to a file. Be sure that your ID variables uniquely identify each observation. I have found that since SAS does all the work for you that your co-workers may not believe that you have validated their listing, so printing the VALID data set gives them a little piece of mind (it is also handy for quickly spotting problems if the comparison results in discrepancies).

```
*****;  
** Compare the data sets, output the results **;  
** and print the validation data set.      **;  
*****;  
  
proc printto file="F:\KENDLE\PHARMSUG\VALID\  
                OUTPUT\VL_DEMO.VAL" new;  
run;  
  
proc compare base=ORIGINAL compare=VALIDATE  
            listbase listcomp;  
  id SITE COL1;  
run;  
  
proc print data=VALIDATE;  
  var SITE COL1-COL10;  
run;  
  
proc printto;  
run;
```

WHAT TO LOOK FOR IN THE PROC COMPARE OUTPUT

By using the LISTBASE and LISTCOMP options, you can double check that the same observations are in both data sets. To be confident that your ORIGINAL and VALIDATE data sets are similar, be sure to check that the number of variables (Nvar) and observations (Nobs) are the same (see the top of the comparison output). It is also a good idea to check that the format of the variables is the same, the output will indicate if the formats are not the same. Finally look for this message at the bottom of the output "NOTE: No unequal values were found. All values compared are exactly equal." When you see this message in addition to matching Nvar and Nobs values then your job of validation is done!

See Appendix 3 for an example of a comparison where the validation DID NOT work.

See Appendix 4 for an example of a comparison where the validation DID work.

Note the items that are bolded in these examples. Those are the portions of the output that need to be examined to determine if the validation was successful or not.

PROS AND CONS

As with any procedure, there are pros and cons of using SAS to validate listings electronically using PROC COMPARE.

PROS

- 100% validation of all the information
- results are concise and problems with the data are easily identified
- human error is minimized by removing the need for hand checking
- process is quick and can be rerun for multiple validations
- uses resources wisely
- electronic results save trees and paper costs

CONS

- output does not look like anything has been done
- lots of extra code is needed if columns have to be "unwrapped"
- format changes to the original output cause SAS code to be modified
- may not meet audit criteria

CONCLUSION

By letting SAS do an electronic validation, you can save time and money in addition to achieving a more accurate validation than could be completed by hand. Electronic validation works best for data listings, but is also an effective way to validate any other output with multiple pages of information that does not cross columns or get wrapped within a column. Hopefully this information will allow you to spend more time programming and less time making manual comparisons.

TRADEMARK INFORMATION

SAS is a registered trademark of SAS Institute Inc.

CONTACT INFORMATION

Lara E.H. Guttadauro
Kendle International Inc.
1200 Carew Tower
441 Vine Street
Cincinnati, OH 45202
Work Phone: (513) 763-1363
Email: guttadauro.lara@kendle.com

Appendix 1: Example of a listing output.

L_DEMO.SAS

17JUL2000

Page 1 of 1

Study Title
Protocol Number

Listing 1
Demographic Characteristics of Randomized Patients

PATIENT ID	PATIENT INITIALS	TREAT	SEX	AGE (a)	DATE OF BIRTH	RACE	HEIGHT (cm) *	WEIGHT (kg) *	STUDY OUTCOME
<<< INVESTIGATIONAL SITE: 001 >>>									
101	DAG	TREATMENT	MALE	31	20JUL1969	AFRICAN AMERICAN	175	86.8	LOST TO FOLLOW-UP
102	IKC	PLACEBO	MALE	49	17DEC1950	CAUCASIAN	170	122.3	COMPLETED STUDY
103	K-Z	TREATMENT	MALE	48	08JAN1952	HISPANIC	178	91.8	WITHDRAWN
104	KAN	PLACEBO	MALE	49	25JUL1951	CAUCASIAN		102.5	LOST TO FOLLOW-UP
105	WAS	TREATMENT	FEMALE	55	03DEC1944	CAUCASIAN	170	78.9	COMPLETED STUDY
106	FMW	PLACEBO	MALE	51	24DEC1948	CAUCASIAN	180	97.0	COMPLETED STUDY
<<< INVESTIGATIONAL SITE: 002 >>>									
201	F-K	TREATMENT	FEMALE	49	10SEP1950	PHILLIPINO	173	80.0	COMPLETED STUDY
202	PAR	TREATMENT	FEMALE	49	24FEB1951	HISPANIC	157	71.8	COMPLETED STUDY
204	SWT	PLACEBO	FEMALE	38	28OCT1961	CAUCASIAN	163	60.2	COMPLETED STUDY
205	NAS	PLACEBO	FEMALE	45	21APR1955	HISPANIC	147	50.1	COMPLETED STUDY
206	KLF	TREATMENT	FEMALE	39	07OCT1960	CAUCASIAN	163	75.5	
207	MAG	PLACEBO	MALE	34	01MAY1966	HISPANIC	160	80.1	WITHDRAWN

(a) Age in years at time of screening.

* At time of screening.

Appendix 2: Proc print of the ORIGINAL data set.

SITE	COL1	COL2	COL3	COL4	COL5	COL6	COL7	COL8	COL9	COL10
001	101	DAG	TREATMENT	MALE	31	20JUL1969	AFRICAN AMERICAN	175	86.8	LOST TO FOLLOW-UP
001	102	IKC	PLACEBO	MALE	49	17DEC1950	CAUCASIAN	170	122.3	COMPLETED STUDY
001	103	K-Z	TREATMENT	MALE	48	08JAN1952	HISPANIC	178	91.8	WITHDRAWN
001	104	KAN	PLACEBO	MALE	49	25JUL1951	CAUCASIAN		102.5	LOST TO FOLLOW-UP
001	105	WAS	TREATMENT	FEMALE	55	03DEC1944	CAUCASIAN	170	78.9	COMPLETED STUDY
001	106	FMW	PLACEBO	MALE	51	24DEC1948	CAUCASIAN	180	97.0	COMPLETED STUDY
002	201	F-K	TREATMENT	FEMALE	49	10SEP1950	PHILLIPINO	173	80.0	COMPLETED STUDY
002	202	PAR	TREATMENT	FEMALE	49	24FEB1951	HISPANIC	157	71.8	COMPLETED STUDY
002	204	SWT	PLACEBO	FEMALE	38	28OCT1961	CAUCASIAN	163	60.2	COMPLETED STUDY
002	205	NAS	PLACEBO	FEMALE	45	21APR1955	HISPANIC	147	50.1	COMPLETED STUDY
002	206	KLF	TREATMENT	FEMALE	39	07OCT1960	CAUCASIAN	163	75.5	
002	207	MAG	PLACEBO	MALE	34	01MAY1966	HISPANIC	160	80.1	WITHDRAWN

**Appendix 3: Example of a comparison where the validation DID NOT work.
Conclusion: Validation Failed**

COMPARE Procedure Comparison of WORK.ORIGINAL with WORK.VALIDATE (Method=EXACT)

```

--- Data Set Summary ---
Dataset              Created              Modified  NVar   NObs
WORK.ORIGINAL      25FEB01:16:02:16  25FEB01:16:02:16   11    12
WORK.VALIDATE      25FEB01:16:02:17  25FEB01:16:02:17   12    11
  
```

```

--- Variables Summary ---
Number of Variables in Common: 11.
Number of Variables in WORK.VALIDATE but not in WORK.ORIGINAL: 1.
Number of ID Variables: 2.
  
```

Listing of Variables in WORK.VALIDATE but not in WORK.ORIGINAL

```

Variable Type Length
CENTER   Char   132
  
```

Comparison Results for Observations
Observation 5 in WORK.ORIGINAL not found in WORK.VALIDATE: SITE=001 COL1=105.

```

--- Observation Summary ---
Observation   Base  Compare  ID
First Obs     1      1  SITE=001 COL1=101
First Unequal 1      1  SITE=001 COL1=101
Last Unequal  12     11  SITE=002 COL1=207
Last Obs      12     11  SITE=002 COL1=207
  
```

```

Number of Observations in Common: 11.
Number of Observations in WORK.ORIGINAL but not in WORK.VALIDATE: 1.
Total Number of Observations Read from WORK.ORIGINAL: 12.
Total Number of Observations Read from WORK.VALIDATE: 11.
  
```

```

Number of Observations with Some Compared Variables Unequal: 11.
Number of Observations with All Compared Variables Equal: 0.
  
```

```

--- Values Comparison Summary ---
Number of Variables Compared with All Observations Equal: 8.
Number of Variables Compared with Some Observations Unequal: 1.
Total Number of Values which Compare Unequal: 11.
  
```

```

Variables with Unequal Values
Variable Type Len Ndif MaxDif
COL9     CHAR  10  11
  
```

Value Comparison Results for Variables

SITE	COL1	Base Value COL9	Compare Value COL9
001	101	86.8	86.80
001	102	122.3	122.30
001	103	91.8	91.80
...

Proc print of the VALIDATE data set.

SITE	CENTER	COL1	COL2	COL3	COL4	COL5	COL6	COL7	COL8	COL9	COL10
001	001	101	DAG	TREATMENT	MALE	31	20JUL1969	AFRICAN AMERICAN	175	86.80	LOST TO FOLLOW-UP
001	001	102	IKC	PLACEBO	MALE	49	17DEC1950	CAUCASIAN	170	122.30	COMPLETED STUDY
001	001	103	K-Z	TREATMENT	MALE	48	08JAN1952	HISPANIC	178	91.80	WITHDRAWN
001	001	104	KAN	PLACEBO	MALE	49	25JUL1951	CAUCASIAN		102.50	LOST TO FOLLOW-UP
001	001	106	FMW	PLACEBO	MALE	51	24DEC1948	CAUCASIAN	180	97.00	COMPLETED STUDY
002	002	201	F-K	TREATMENT	FEMALE	49	10SEP1950	PHILLIPINO	173	80.00	COMPLETED STUDY
002	002	202	PAR	TREATMENT	FEMALE	49	24FEB1951	HISPANIC	157	71.80	COMPLETED STUDY
002	002	204	SWT	PLACEBO	FEMALE	38	28OCT1961	CAUCASIAN	163	60.20	COMPLETED STUDY
002	002	205	NAS	PLACEBO	FEMALE	45	21APR1955	HISPANIC	147	50.10	COMPLETED STUDY
002	002	206	KLF	TREATMENT	FEMALE	39	07OCT1960	CAUCASIAN	163	75.50	
002	002	207	MAG	PLACEBO	MALE	34	01MAY1966	HISPANIC	160	80.10	WITHDRAWN

Appendix 4: Example of a comparison where the validation DID work. Conclusion: Validation Successful

COMPARE Procedure Comparison of WORK.ORIGINAL with WORK.VALIDATE (Method=EXACT)

```

--- Data Set Summary ---
Dataset              Created              Modified  NVar   NObs
WORK.ORIGINAL      25FEB01:15:31:40  25FEB01:15:31:40   11    12
WORK.VALIDATE      25FEB01:15:31:40  25FEB01:15:31:40   11    12
  
```

```

--- Variables Summary ---
Number of Variables in Common: 11.
Number of ID Variables: 2.
  
```

```

--- Observation Summary ---
Observation          Base Compare ID
First Obs            1           1  SITE=001 COL1=101
Last Obs             12          12  SITE=002 COL1=207
  
```

Number of Observations in Common: 12.
 Total Number of Observations Read from WORK.ORIGINAL: 12.
 Total Number of Observations Read from WORK.VALIDATE: 12.

Number of Observations with Some Compared Variables Unequal: 0.
 Number of Observations with All Compared Variables Equal: 12.

NOTE: No unequal values were found. All values compared are exactly equal.

Proc print of the VALIDATE data set.

SITE	COL1	COL2	COL3	COL4	COL5	COL6	COL7	COL8	COL9	COL10
001	101	DAG	TREATMENT	MALE	31	20JUL1969	AFRICAN AMERICAN	175	86.8	LOST TO FOLLOW-UP
001	102	IKC	PLACEBO	MALE	49	17DEC1950	CAUCASIAN	170	122.3	COMPLETED STUDY
001	103	K-Z	TREATMENT	MALE	48	08JAN1952	HISPANIC	178	91.8	WITHDRAWN
001	104	KAN	PLACEBO	MALE	49	25JUL1951	CAUCASIAN		102.5	LOST TO FOLLOW-UP
001	105	WAS	TREATMENT	FEMALE	55	03DEC1944	CAUCASIAN	170	78.9	COMPLETED STUDY
001	106	FMW	PLACEBO	MALE	51	24DEC1948	CAUCASIAN	180	97.0	COMPLETED STUDY
002	201	F-K	TREATMENT	FEMALE	49	10SEP1950	PHILLIPINO	173	80.0	COMPLETED STUDY
002	202	PAR	TREATMENT	FEMALE	49	24FEB1951	HISPANIC	157	71.8	COMPLETED STUDY
002	204	SWT	PLACEBO	FEMALE	38	28OCT1961	CAUCASIAN	163	60.2	COMPLETED STUDY
002	205	NAS	PLACEBO	FEMALE	45	21APR1955	HISPANIC	147	50.1	COMPLETED STUDY
002	206	KLF	TREATMENT	FEMALE	39	07OCT1960	CAUCASIAN	163	75.5	
002	207	MAG	PLACEBO	MALE	34	01MAY1966	HISPANIC	160	80.1	WITHDRAWN

