

Paper CC-08

Be Bold with Proc Compare and %RTFTable

Patricia Guldin, Merck Research Labs, Merck & Co., Inc., Upper Gwynedd, PA

ABSTRACT

Comparing data sets and reviewing data are quite common tasks. New data comes in, data points get updated and people need and want to know about it. When reviewing new or modified data, it is often desirable to see all of the data. This often results in repeatedly running the same programs, looking for changes in ever growing reports. What an achievement it would be if the new and modified data popped out at us; life would be easier and programmers would become heroes in the eyes of those who review the data. This paper will spark ideas of ways to make this happen using the stylevar parameter of %RTFTable in conjunction with PROC COMPARE and will provide a specific example of one use of this technique.

INTRODUCTION

Do you run programs that produce tables for data integrity review, medical monitoring, external or internal committee review? Do you produce tables each time new data is added or data is modified, or at specified time points to see new or modified data? Have you ever heard "Can you give me tables that show just the new and updated information?" and seconds later, "But still include all of the data so I can have a better reference point"? You may have used PROC COMPARE to see the differences, but if you add a style variable with %RTFTable, you can see the differences as bold, italics, or underlined.

There are two fairly basic steps to achieve this, identify the changed data and report the data.

Step one is to identify what you want to highlight by flagging new and/or changed records in data sets.

Step two is to use a reporting macro to output the data, highlighting the flagged records.

STEP ONE

To begin, save the final data set each time you create your table. Define the locations of the old and new data sets. Sort the two data sets using the same variables in the same order. Using PROC COMPARE, compare the old and new data sets and add flags to new and/or changed records in the newest data set.

PROC COMPARE compares two SAS® data sets. It can identify matching variables and matching observations, and thus identify non-matching or new observations. It helps to know a little about PROC COMPARE so we can use it to create our tables.

The basic syntax for PROC COMPARE needed for this paper is:

```
PROC COMPARE<BASE=dset> <COMPARE=dset>
  OUT= dscomp OUTNOEQUAL OUTCOMP NOPRINT;
  ID variable(s)
  VAR variable(s);
RUN;
```

The base data set in this case contains the original, old, or previous version of the data. The compare data set in this case contains the new or updated data. The compare data set in this example is the one where we expect to see differences or changes.

OUT= creates an output data set. DSCOMP is the name of the output data set.

OUTNOEQUAL, OUTCOMP, and NOPRINT are options.

The OUTNOEQUAL option will suppress an observation being written to our output data set, dscomp, when all values in the observation are equal. This produces a data set containing only the unequal observations.

The OUTCOMP option writes observations to the output data set and creates observations in which `_TYPE_ = COMPARE`. Using this option will allow us to pull out the new and changed records so they can be flagged.

OUTBASE would create a record in the output data set for each record in the base data set that didn't change and would assign `_TYPE_ = BASE`. OUTDIF would create a record with `_TYPE_ = DIF` in the output data set only for the records that exist in the base and compare data sets but have differences in the specified variables. For this example using OUTCOMP is the best option as it allows us to identify new and changed records in one step. Depending on your need you could use any combination of these options with merging data sets to get what you want out of the compare.

The option NOPRINT suppresses the printing of the output data set.

ID is used to identify the variable(s) used to match observations. Both data sets must be sorted by the ID variable(s). This is how we know we are comparing the same observations in each data set. For example you might want to match observations on subject id, event name and event date.

VAR restricts the comparison to values of the variables listed. This is where we would check for changes in an observation. For example, for matched observations, look at severity of the event and stop date of event for changes in these variables from the prior version of the data set. By not specifying VAR we can compare all variables. This is the option chosen for the example described in this paper.

STEP TWO

%RTFTable is a SAS macro that was developed at Merck & Company in 1995 to generate rich text format (RTF) tables. It has been used since then with minimal upgrades. This macro controls table appearance in the title, header, body and footnotes. This is the reporting macro that I used for this example because it is the reporting macro most commonly used by my company. Any reporting macro with the ability to control font formatting could have been used. I have used the stylevar parameter of the %RTFTable macro to make the flagged records appear differently in the output table.

The focus of this paper is on controlling the appearance of the rows in the body of the table using the stylevar parameter of the %RTFTable macro. This parameter will allow certain rows to stand out from the rest of the rows in a table. The stylevar parameter applies specific style formats for three areas, font, horizontal line, and vertical line formatting. The example in this paper uses the font formatting.

To use the stylevar parameter you must create a variable, i.e. STYLE. You must assign STYLE the value of B for bold, I for italic, UL for underline, or P or blank for plain in the data set. The style of the rows in the table corresponding to these records will be the value of STYLE. One limitation is that you can only select one value for the value of STYLE. For example, you cannot use both bold and italic; you must choose one or the other.

Another minor drawback is that the style variable applies the selected style at the row level. It cannot be applied at the cell level. This means that even if only one variable has a change, the entire row in the output will be in bold. You will see this in the output below. For some rows only one variable changed, but the entire row is bold, so you can not distinguish exactly what the change was from this single report. You could change the output and display the original record and the new record but in some situations that may cause the report to become unwieldy.

SPECIFIC EXAMPLE

This example uses adverse event data sets. In this example some variables were used to match observations (subject, ae term, and ae start date) and some variables were used in the compare to determine new or modified records (subject, ae term, ae start date, severity, or relationship.)

It is not required to use all variables for both. You can match using one set of variables and compare the records based on another set of variables. Typically there would be some overlap in the variables specified in the ID and VAR statements. If the VAR statement only included aeseverity and not aerelation and if only the drug relationship changed, the record would not be flagged and will not appear differently from the unchanged rows in the output table. If the severity changes, the record would be flagged and appear differently from unchanged rows in the output table.

Bold is used as the style. The data sets and sample code are below. Bold italics are used in the AENEW data set to show the changes and new data that should appear bolded in the output table.

DATA SETS

AELAST – BASE DATASET

<u>AESEVERITY</u>	<u>AERELATION</u>	<u>AESTDT</u>	<u>AETERM</u>	<u>SUBJID</u>
MILD	NOT RELATED	2008-06-18	Depression	13069
MILD	NOT RELATED	2008-01-30	Ear neoplasm malignant	13069
MILD	NOT RELATED	2008-04-14	Flatulence	13069
UNKNOWN	NOT RELATED	2008-02-08	Injection site swelling	13069
MILD	NOT RELATED	2008-04-14	Joint swelling	13069
MILD	NOT RELATED	2008-04-14	Sinus headache	13069
MODERATE	NOT RELATED	2008-08-10	Depression	13061
MODERATE	NOT RELATED	2008-10-08	Excoriation	13061
MILD		2008-12-18	Agitation	15100
MODERATE	RELATED	2008-12-16	Injection site pain	15100
MILD	RELATED	2008-12-17	Injection site swelling	15100
MILD	NOT RELATED	2008-12-25	Stomatitis	15100

AENEW – COMPARE DATA SET

<u>AESEVERITY</u>	<u>AERELATION</u>	<u>AESTDT</u>	<u>AETERM</u>	<u>SUBJID</u>
MILD	NOT RELATED	2008-07-19	Depression	13069
MILD	NOT RELATED	2008-01-23	Ear neoplasm malignant	13069
MILD	NOT RELATED	2008-04-15	Epistaxis	13069
MILD	NOT RELATED	2008-04-14	Flatulence	13069
MILD	NOT RELATED	2008-04-14	Hyperchlorhydria	13069
UNKNOWN	RELATED	2008-02-08	Injection site swelling	13069
MILD	NOT RELATED	2009-01-15	Intention tremor	13069
MILD	NOT RELATED	2008-04-14	Joint swelling	13069
MILD	NOT RELATED	2008-04-14	Myalgia	13069
MILD	NOT RELATED	2009-02-03	Neurological examination abnormal	13069
MILD	NOT RELATED	2008-04-14	Sinus headache	13069
MODERATE	NOT RELATED	2008-08-10	Depression	13061
MODERATE	NOT RELATED	2008-10-08	Excoriation	13061
MILD	NOT RELATED	2009-04-28	Anxiety	13214
MODERATE	RELATED	2008-12-16	Injection site pain	15100
MILD	RELATED	2009-02-16	Injection site pain	15100
UNKNOWN	RELATED	2008-12-17	Injection site swelling	15100
UNKNOWN	RELATED	2009-02-19	Injection site swelling	15100
MILD	RELATED	2008-12-18	Mental status changes	15100
MILD	NOT RELATED	2009-02-17	Sinus congestion	15100
MILD	NOT RELATED	2009-03-02	Sinus congestion	15100
MILD	NOT RELATED	2008-12-25	Stomatitis	15100

SAMPLE CODE

```

/*  STEP ONE  */

/*  Sort data sets  */

PROC SORT data=aelast;    *original data set*;
  BY subjid aeterm aestdt aeseverity aerelation;
RUN;

PROC SORT data=aenew;    *newest data set*;
  BY subjid aeterm aestdt aeseverity aerelation;
RUN;

/*  Compare data sets  */

PROC COMPARE base=aelast compare=aenew
  OUT= dscomp OUTNOEQUAL OUTCOMP NOPRINT;
  ID subjid aeterm aestdt;
RUN;

/*  Flag new and/or modified records  */

%MACRO flagae();
DATA _NULL_;
  IF 0 THEN SET dscomp NOBS=total;
  CALL SYMPUT('ndiff',left(put(total,8.)));
  STOP;
RUN;

%PUT ndiff=&ndiff;

%IF &ndiff >0 %THEN %DO;
  DATA aelst(drop=_type_ _obs_);
    MERGE aenew(in=a)
          dscomp(in=b where (_type_="COMPARE"));
  BY subjid aeterm aestdt aeseverity aerelation;
  LENGTH modflag $1.;
  IF b THEN DO;
    modflag='Y';
  END;
  RUN;
%END;

%ELSE %DO;
  DATA aelst;
    SET aenew;
    BY subjid aeterm aestdt aeseverity aerelation;
    modflag = " ";
  RUN;
%END;

%MEND flagae;

%flagae();

```

```

/* Set the STYLE variable to bold */

DATA aelst;
  SET aelst;
  BY subjid aeterm aestdt aeseverity aerelation;
  IF modflag='Y' THEN STYLE = 'B';
RUN;

/* STEP TWO */

/* Create output table */

%LET title1= Listing of Adverse Event Data with New/Modified Data in Bold;
%LET header=1|Subject|AE Term|Start Date|Severity|Drug Relationship|;

%RTFTable(
  DS           = aelst
  ,TITLEROW    = 1|12|B
  ,HEADROWS    = 1|8|B
  ,STYLEVAR    = STYLE|none|same|
  ,NOREPEAT    =
  ,SKIPBY      =
  ,PAGEBY      =
  ,VAR         = subjid aeterm aestdt aeseverity aerelation
  ,FMT         = $6. $200. $20. $15. $15.
  ,WDTH        = 10 40 15 20 15
  ,ALIGN       = C C C C C
  ,BORDERS     = |||||
  ,DOCLIBR     = &outdir
  ,DOCFILE     = aelist.rtf
  ,FONTSIZE    = 10
  ,PORTLAND    = P
);

```

FIGURE 1. SAMPLE OUTPUT TABLE.

Listing of Adverse Event Data with New/Modified Data in Bold

Subject	AE Term	Start Date	Severity	Drug Relationship
13061	Depression	2008-08-10	MODERATE	NOT RELATED
13061	Excoriation	2008-10-08	MODERATE	NOT RELATED
13069	Depression	2008-07-19	MILD	NOT RELATED
13069	Ear neoplasm malignant	2008-01-23	MILD	NOT RELATED
13069	Epistaxis	2008-04-15	MILD	NOT RELATED
13069	Flatulence	2008-04-14	MILD	NOT RELATED
13069	Hyperchlorhydria	2008-04-14	MILD	NOT RELATED
13069	Injection site swelling	2008-02-08	UNKNOWN	RELATED
13069	Intention tremor	2009-01-15	MILD	NOT RELATED
13069	Joint swelling	2008-04-14	MILD	NOT RELATED
13069	Myalgia	2008-04-14	MILD	NOT RELATED
13069	Neurological examination abnormal	2009-02-03	MILD	NOT RELATED
13069	Sinus headache	2008-04-14	MILD	NOT RELATED
13214	Anxiety	2009-04-28	MILD	NOT RELATED
15100	Injection site pain	2008-12-16	MODERATE	RELATED
15100	Injection site pain	2009-02-16	MILD	RELATED
15100	Injection site swelling	2008-12-17	UNKNOWN	RELATED
15100	Injection site swelling	2009-02-19	UNKNOWN	RELATED
15100	Mental status changes	2008-12-18	MILD	RELATED
15100	Sinus congestion	2009-02-17	MILD	NOT RELATED
15100	Sinus congestion	2009-03-02	MILD	NOT RELATED
15100	Stomatitis	2008-12-25	MILD	NOT RELATED

CONCLUSION

This paper provides an example of how to combine PROC COMPARE with %RTFTable macro and the stylevar parameter to enhance output. This provides a way to display cumulative study data in a format that allows customers to easily recognize where there is new/changed data. This helped make regular review of data quicker and easier. There are varied uses for this technique. Give it a try!

REFERENCES

Casas, Angelina Cecelia. "Proc Compare to Validate Datasets".
<http://www.lexjansen.com/pharmasug/2003/tutorials/tu056.pdf>.

Qi, Eric and Zhang, Liping. "%RTFTable – A Powerful SAS® Tool to Produce Rich Text Format Tables".
<http://www.lexjansen.com/pharmasug/2003/technicaltechniques/tt038.pdf>.

Smith, Jared. "When is New Really New? Proc Compare Gives the Details".
<http://www.nesug.org/Proceedings/nesug06/dm/da28.pdf>.

SAS Institute Inc. (2003), Online User Documentation, Cary, NC: SAS Institute Inc.
<http://support.sas.com/onlinedoc/913/docMainpage.jsp>

ACKNOWLEDGMENTS

I would like to thank Ms. Liping Zhang for reviewing the paper and encouraging me to write it.

CONTACT INFORMATION

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

Patricia Guldin
Merck Sharp & Dohme Corp.
UG 1D-10
Upper Gwynedd, PA 19454
(267) 305-8242
patricia_guldin@merck.com

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.