Paper CC26

A SAS Macro to Automate the QC of CRT SAS Transport Files (using ODS 'Traffic Lighting')

Todd Case, Genentech, Inc., South San Francisco, CA
Ed Stamm, DataCeutics, Inc., Pottstown, PA

ABSTRACT

In the past, SAS® Programmers may have had to manually Quality Control ('QC') each CRT (Case Report Tabulation) SAS Transport file before it was submitted to the FDA. This paper introduces a SAS macro which aids in the speedy and efficient QC of the CRT SAS Transport files headed to the FDA - across therapeutic areas and different drugs. Specifically, this program facilitates visual QC of the CRT SAS Transport files by:

(1) Generating a list of the CRT dataset variables and variable attributes (in the order they appear in the CRT datasets) with problematic variables highlighted using ODS 'Traffic Lighting'. The footnotes in the macro output provide the criteria for 'problematic variables', including variables with formats other than date/time formats, variables with informats, variables with names greater than 8 characters, labels greater than 40 characters, etc. as well as,

(2) Generating QC SAS Transport files from the source study(s) and comparing them (via 'Proc Compare') to the CRT SAS transport files headed to the FDA.

INTRODUCTION

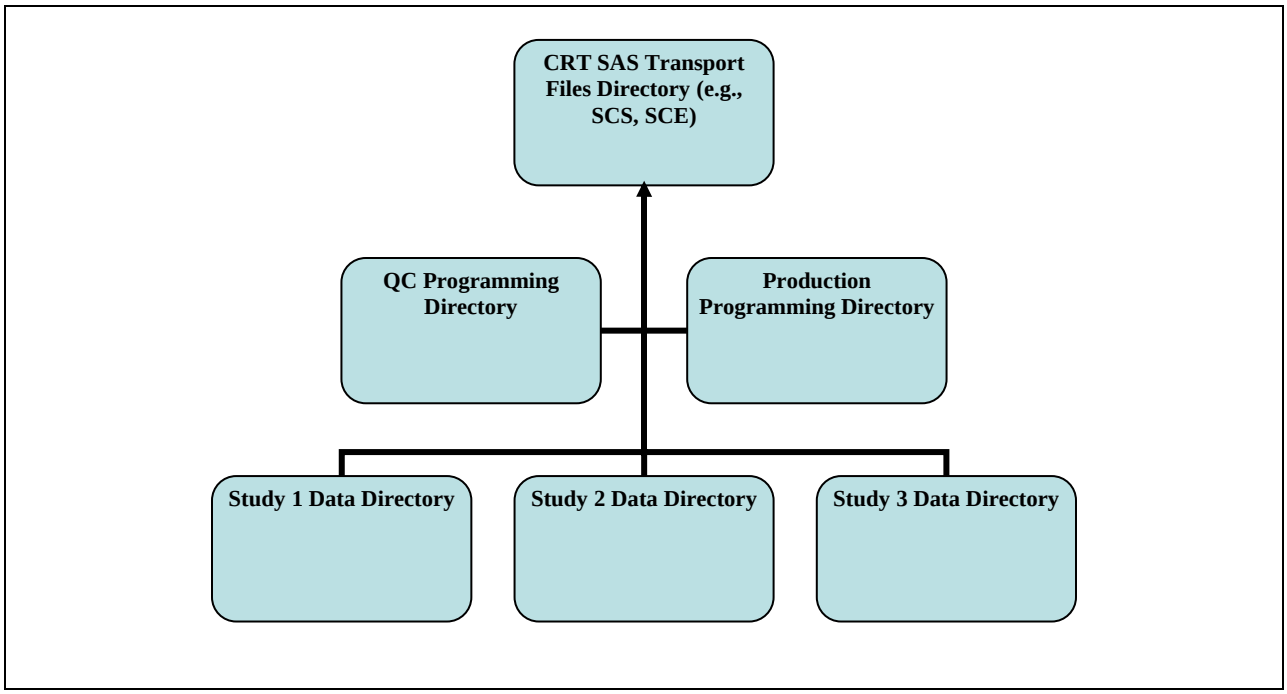
Before CRT study datasets are submitted to the FDA, they often need to be manipulated in various ways (variable/dataset attributes may need to be modified, etc.) and subsequently converted into SAS Transport files. In turn, these SAS Transport files require QC to ensure data integrity was not compromised during the conversion process. For the purposes of this paper QC is defined as a "Product oriented checks to ensure that results are accurate and reliable".

This paper will discuss the following: (1) the rationale for creating the macro, (2) the macro assumptions and parameters, and (3) the output produced by the macro.

RATIONALE

In many instances, the time immediately preceding sending SAS Transport files (generated by the CRT datasets) to the FDA can be fraught with stress and headaches for SAS Programmers, who may need to change dataset and variable labels or other attributes at the time when accuracy is most critical. Regardless of this dilemma, it is ultimately the responsibility of the SAS Programmer to ensure these SAS Transport files are accurate and meet the specifications and attributes decided on by the clinical team (as well as the FDA!). The macro described in this paper is especially useful when time is of the essence and datasets going to the FDA require a formal QC (e.g., some of these situations may be in the form of a BLA/NDA, ISS/SCS, etc.). Figure 1 (below) represents a possible schema of a directory containing integrated data which requires QC - and hence an ideal situation when the macro described in this paper could be useful.

Figure 1
Possible SCS / SCE Schema



MACRO ASSUMPTIONS AND PARAMETERS

This macro has three assumptions and seven parameters.

Assumptions:

(1) CRT datasets and their corresponding SAS transport files reside in the same directory (e.g., ae.sas7bdat and ae.xpt are in the same directory),
(2) CRT datasets are sorted (QC program sorts datasets by the same variables as the Developer program),
(3) SAS9 is used.

Parameters:

(1) BASEDIR (required): Directory where QC datasets will be generated *from*
(2) COMPDIR (required): Directory holding CRT datasets which QC datasets will be *compared to*
(3) DSETS (required): List of datasets (in BASEDIR directory) to be *included* in QC
(4) DROPVARS: List of variables to be dropped – if populated, these must exist in every dataset
(5) RENAMVRS: List of variables to be renamed – if populated, these must exist in every dataset
(6) VARORD: Sort order of CRT variables - these must exist in every dataset
(7) KEEPXPT: Keep QC SAS transport files?

MACRO OUTPUT

This macro generates a single html file (v_d_esub.html) with two outputs for each dataset, as follows:

(1) A list of variables and variable attributes in the order they appear in the CRT dataset (Figure 3 – with problematic variables are highlighted in red), and
(2) A comparison of the CRT SAS transport file generated by the Developer vs. the SAS transport file generated by the QC Programmer (Figure 4).

As you can see if Figure 3 (the 1st of the two outputs mentioned above), there are 6 titles which give information such as the number of datasets being QC'd, the full pathname of the CRT dataset, as well as the dataset label and length and the variables the CRT dataset is sorted by. In addition, the main body of the output lists every single variable in the dataset (with 'problematic' variables highlighted in red, using ODS 'Traffic Lighting' within the Proc Report procedure). Lastly, the ten footnotes display all the criteria that will cause a variable to be highlighted in red (e.g., if any of these criteria are met the row will be highlighted in red).

Figure 3
SAS HTML Output Example using ODS 'Traffic Lighting'

CRT Dataset (1 of 1): */fullunixpathname/*.ae
CRT File Size (bytes): 2564096
CRT Dataset label: Adverse Events
CRT Dataset label length: 14
CRT Dataset is sorted by these variables: STUDY PATNUM
A row in red indicates a problematic variable (e.g., it meets >= 1 criteria specified in the footnotes).

Variable Name	Variable Name Length	Correct Variable Name	Variable Number	Variable Format	Variable Informat	Variable Type	Variable Length	Maximum Variable Length	Variable Label	Variable Label Length
PATNUM	6	PATNUM	1			1			Subject	7
STUDY	5	STUDY	2			2	6	5	Study Number	12
AGE	3	AGE	3			1			Age (years) at Randomization	28
RACE	4	RACE	4			2	16	16	Race	4
SEX	3	SEX	5			2	6	6	Sex	3
…	…	…	…	…	…	…	…	…	…	…

Ensure the dataset label length (Title 4) is <= 40 characters.
Ensure sorted by variables (Title 5) match those specified in specs.
Ensure variable names (Column 1) do not begin with an underscore.
Ensure Column 2 (Variable Name Length) is not > 8.
Ensure Column 3 (Correct Variable Name) = Column 1 (Variable Name) - if not CRT dataset needs to reorder variables.
Ensure there are no formats other than DATE/TIME/DATETIME and that there are no informats.
Ensure character variables are left aligned.
Ensure character variable lengths (Column 8) are not longer than 200 characters.
Ensure Column 8 (Variable Length) is not larger than Column 9 (Maximum Variable Length).
Ensure Column 11 (Variable Label Length) is not > 40.

Please note the following in Figure 4 (the 2nd of the two outputs as described above). The output shows that the QC SAS Transport file is being compared to the SAS Transport file going to the FDA (in the example below we can see that the QC dataset doesn't have a label, in this case it is confirmed that the dataset being sent to the FDA has the correct label). We can also see that two variables have different attributes but for propriety reasons we will only note that the label was different and the labels going to the FDA were accurate and appropriate.

Figure 4
Proc Compare Output (Abbreviated)

```
The COMPARE Procedure
Comparison of DEV.AE with QC.AE
(Method=EXACT)

Data Set Summary

Dataset            Created           Modified  NVar     NObs  Label

DEV.AE    13MAR06:11:05:44  13MAR06:11:05:44    31        .  Adverse Events
QC.AE     05FEB07:14:47:28  05FEB07:14:47:28    31        .


Variables Summary

Number of Variables in Common: 31.
Number of Variables with Differing Attributes: 29.
```

…

```
Observation Summary

Observation      Base  Compare

First Obs           1        1
Last  Obs        2774     2774

Number of Observations in Common: 2774.
Total Number of Observations Read from DEV.AE: 2774.
Total Number of Observations Read from QC.AE: 2774.

Number of Observations with Some Compared Variables Unequal: 0.
Number of Observations with All Compared Variables Equal: 2774.

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

CONCLUSION

Employing a macro (v_d_esub.sas) to generate and manipulate QC SAS Transport files automates what programmers may have previously been performing manually. A major benefit of using a single macro to QC these SAS Transport files is that individual programmers do not need to copy their programs from directory to directory. This macro can either be called from a macro library or, at a minimum, be maintained in a macro library.

Even if this macro is called with only the required parameters it has inherent value in that it employs a very conservative approach - it will just generate XPT files based on the original source data and User can visually compare the differences that were made during the conversion process. This can be very useful when a thorough QC is necessary and time is of the essence - which is always the case in our industry!

ACKNOWLEDGMENTS

Thanks to Sarbjit Rai for initiating this project, Curtis Wolf and Sharon Hall for technical input, and Sandra Minjoe for reviewing this paper.

RECOMMENDED READING

Carpenter, Art. 2004. Carpenter's Complete Guide to the SAS Macro Language, Second Edition. Cary, NC: SAS Institute, Inc.

SAS Support Website, FAQ #3257 "In ODS HTML, can I specify the color and the font of a cell based on the cell's value?". http://support.sas.com/faq/032/FAQ03257.html

CONTACT INFORMATION

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

Todd Case
Genentech, Inc.
MS 84
South San Francisco, CA 94108
Work Phone: 650 467 2410
E-mail: caset@gene.com

Ed Stamm
DataCeutics, Inc.
1610 Medical Drive
Pottstown, Pa 19464
(610) 970-2333
stamme@dataceutics.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.