Paper PP02

An Efficient Way of Managing the Execution and QC of SAS Programs in Multiple Folders

Danny Hsu, Wei Qian; Pfizer Inc., Bothell, WA, USA

ABSTRACT

A clinical trial could involve multiple statistical programmers working on hundreds of SAS programs in different subfolders (e.g., sdtm, adam, and tlfs). It is therefore critical that a lead programmer manages execution and QC efficiently and effectively. This paper proposes a method of setting up a batch file which first executes all programs in proper sequence and then runs a utility program to generate an overall summary report listing file names and authors for programs with unmatched results or unwanted messages (e.g., "ERROR" and "WARNING") in log files. In this way, with one click, hundreds of SAS programs are executed and QCed and the lead programmer gets immediate insight into the study status by utilizing the QC report and assignment information to notify programmers to correct their programs for any issues. This process would streamline the program execution process while continuously tracking QC status.

INTRODUCTION

In many organizations, it is common that a study team has multiple statistical programmers. A typical statistical programming deliverable involves a sequence of production and QC of SDTM, ADaM, and tables, listings, and figures (TLFs). The SAS® or R programs for each of these are typically stored in their respective directory for ease of code management. Within each of these folders, programs also need to be executed in a certain order to ensure proper dependency. Therefore, for a programming lead, it can be a very time-consuming and tedious task to batch run all the programs for a deliverable to ensure proper dependency, as well as have insight into status of each of the programs, such as

- Did the program finish without errors, warnings, or other unwanted message in log files.
- Did the output pass QC.

CREATE BATCH FILE TO RUN PROGRAMS FROM MULTIPLE FOLDERS

Traditionally, to refresh everything within a clinical trial analysis, the lead programmer may first use a batch file to run all SDTM productive programs and wait until the run is done, then use another batch file to run all SDTM QC programs. After that, other batch files would be used to run ADaM production, ADaM QC, TLF production, and finally TLF QC. This process requires the lead programmer to stand by and check the run status and execute the different batch files.

Therefore, it would be beneficial to create a batch file (example in Table 1) which could run all these statistical programs in different folders in a predefined sequence, and the lead programmer could also receive an email notice once the running is done. In this way, the lead programmer does not need to stand by to check run status multiple times, and the batch file could even be set up to run automatically at midnight to streamline the run-all process. Lastly, having such a batch file means the execution process can be reliably repeated, for the lead as well as the back-up when the lead is not available.

Table 1 – Example of batch file

```
:: run SDTM Prod
"C:\SAS\RunSAS.exe" ^ "C:\STUDY\sdtm\prod\batch_run.csv" ^ "%~n0"

:: run SDTM QC
"C:\SAS\RunSAS.exe" ^ "C:\STUDY\sdtm\qc\batch_run.csv" ^ "%~n0"

:: run ADaM Prod
"C:\SAS\RunSAS.exe" ^ "C:\STUDY\adam\prod\batch_run.csv" ^ "%~n0"

:: run ADaM QC
"C:\SAS\RunSAS.exe" ^ "C:\STUDY\adam\qc\batch_run.csv" ^ "%~n0"

:: run TLFs Prod
"C:\SAS\RunSAS.exe" ^ "C:\STUDY\tlfs\prod\batch_run.csv" ^ "%~n0"

:: run TLF QC
"C:\SAS\RunSAS.exe" ^ "C:\STUDY\tlfs\qc\batch_run.csv" ^ "%~n0"

:: check QC results
"C:\SAS\RunSAS.exe" ^ "C:\STUDY\tool\search_unmatched_unwanted_log.csv" ^ "%~n0"
```

CHECK FOR ANY UNMATCHED DOMAINS

PROC COMPARE is a powerful procedure frequently used in data set and output QC, particularly in automated processes. By employing the method described in Deciphering PROC COMPARE Codes: The Use of the bAND Function (Hinson & Coughlin, 2012), the details of PROC COMPARE results could be added into the log files of all QC programs. Once all log files are generated from the previous run-all process, the last row of the batch file in **Table 1** executes a utility program to check all those PROC COMPARE results, find out which domains have mismatched results, and generate a summary report as per **Table 2** to list those unmatched domains with their designated programmers.

Table 2 – List of unmatched domains

FOLDER	FILE	PRIMARY PROGRAMMER	VALIDATOR
sdtm\qc	v-ex.log	Kevin	Ankit
sdtm\qc	v-sv.log	Joyce	Mia
adam\qc	v-adrs.log	Ankit	Tom
adam\qc	v-adsl.log	Joyce	Tom
tlf\qc	v-l-ex.log	Mia	Joyce
tlf\qc	v-t-dm.log	Kevin	Ankit

CHECK FOR ANY UNWANTED LOG MESSAGES

Furthermore, the batch run script would also check the log files and list the ones with unwanted messages (e.g., "ERROR", "WARNING", or "UNINITIALIZED") in the same summary report (e.g., **Table 3**). In this way, the study lead could easily refer to this summary report to track the QC status and clean the log messages.

Table 3 – List of logs with unwanted messages

FOLDER	FILE	PRIMARY PROGRAMMER	VALIDATOR
sdtm\prod	ae.log	Mia	Joyce
sdtm\prod	cm.log	Kevin	Ankit
sdtm\qc	v-cm.log	Kevin	Ankit
adam\prod	adae.log	Mia	Joyce
adam\qc	v-adae.log	Mia	Joyce

CONCLUSION

The whole process above would streamline the program execution process while continuously tracking QC status and log issues. It would definitely improve the efficiency of managing the execution and QC of statistical programs in multiple folders within a study, allow for automation of deliverable execution, and provide run status of all program executions, with a high-level summary of issues and diagnoses that enable the study lead to triage and assign further tasks accordingly.

REFERENCE

Hinson, J. and Margaret, C (2012) "Deciphering PROC COMPARE Codes: The Use of the bAND Function", SAS Global Forum 2012. (https://support.sas.com/resources/papers/proceedings12/063-2012.pdf)

ACKNOWLEDGMENTS

We would like to acknowledge Michael Hagendoorn for his valuable feedback and constant support and guidance.

CONTACT INFORMATION

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

Author Name: Danny Hsu; Wei Qian
Company: Pfizer Inc.
Address: 21717 30th Dr SE, Bothell, WA 98021
Email: danny.hsu@pfizer.com; wei.qian@pfizer.com
Website: https://www.pfizer.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.

Brand and product names are trademarks of their respective companies.