

Building a Clinical SAS® Programming Group from the Ground Up

Devin J. Hunt, Clinical Research Management, Inc. contractor to USAMMDA, Division of Regulatory Affairs and Compliance, Fort Detrick, MD

Nicole C. Close, Clinical Research Management, Inc. contractor to USAMMDA, Division of Regulatory Affairs and Compliance, Fort Detrick, MD

DISCLAIMER

The views expressed in this paper are those of the authors and do not reflect official policy or position of the Department of the Army, Department of Defense or the U.S. Government.

ABSTRACT

Structuring and building a clinical SAS® Programming group from scratch can be a formidable task for just about any group or company. When an organization decides to provide SAS® programming support in-house instead of contracting out, or to supplement programming services, there is an inordinate amount of planning and development that must be initially conducted in order to provide the most effective and streamlined start-up. The United States Food and Drug Administration (FDA) and the International Conference on Harmonization (ICH) have clinical trial reporting and submission guidances and standards that organizations should follow and take into consideration when developing programming policies and procedures. Presented in this paper is one example of how the United States Army Medical Materiel Development Activity (USAMMDA) began the SAS® Programming group within their Biostatistics Branch. Included in the paper will be examples of SAS® Programming Standard Operating Procedures (SOPs), SAS® macro library development and programs used to streamline programming tasks, a model of file directory structures, and steps to validate your SAS® working environment.

KEYWORDS

programming group, validation, SOP, macro library, directory structure, license type

INTRODUCTION

For organizations that are small in size and scope, SAS® programming is typically outsourced to other companies and there resides someone in the sponsor organization to manage the hired contracting company. There are many reasons as to why SAS® programming is outsourced with some being a lack of knowledge about requirements, inadequate personnel or staff, limited computer resources, or there is insufficient need within the organization to justify the costs and resources. However, as organizations grow in size, some decide to build an in-house programming group and there are many considerations to take into account prior to making a final decision. For the United States Army Medical Materiel Development Activity (USAMMDA) within the US Army Medical Research and Materiel Command (USAMRMC), the decision to start the Biostatistics Branch and provide SAS® programming support stemmed from a desire to provide a full range of regulatory support to the Command for Biostatistics, rely less on outside contracting organizations, and to have subject matter experts to oversee existing and new contracting companies. USAMMDA also wanted to reduce the timeframe from protocol development to study start-up. When starting a new group within an organization, there are many organization specific rules and policies to follow. This paper outlines the steps the USAMMDA Biostatistics Branch took in order to develop the SAS® programming support the Command needed.

MAIN IDEA

ACQUIRING SAS®

The first task in starting a SAS® programming group is to determine what SAS® packages and the number of licenses you will need for your group. You will also want to take into consideration the type of licenses you will need. You can decide from SAS® server licenses or local PC licenses. There are advantages and disadvantages to each type of license and the type you select will depend upon the needs of your organization.

Server licenses and environments are typically cost prohibitive for smaller companies. Along with purchasing the software and licenses, you will also need to have a server to run your SAS® software. Usually having a SAS® dedicated server is the best set up, but this will depend on your network and SAS® programming demands. With a server also comes the technical support needed to setup and maintain the server. Though this option is a bit most costly than a local license, there are advantages to this set up which will pay off when your programming group grows

in size. The SAS® server environment allows for easier deployment of upgrades, updates, and hotfixes. The server environment also ensures that all programmers are using the exact same validated SAS® environment.

For most small organizations, local PC licenses will probably be the most cost effective solution. These licenses are assigned on a machine basis and each computer SAS® is installed on will need its own license. Since the SAS® software is installed directly on the local machine, it does not need to be connected to the network in order to run SAS®. There is no need for a server or the technical support to set up and maintain the SAS® server. The local licenses are also less expensive than the server licenses. The main disadvantage to this approach is the loss of control over the machine's environment. The user may have the ability to change or install applications that could adversely affect the SAS® software once it has been validated. Stricter controls will need to be implemented in order to ensure the SAS® system remains in a validated state and the correct updates and hotfixes are applied to all machines.

Once you decided on the type of environment SAS® will be running in, the next step is to choose what types of packages to buy. The types of SAS® packages and add-ons will be another area you will need to research. The types of packages and add-ons will again depend on the need of your organization. Some questions to ask when deciding the types of packages to buy should include "Will our organization be conducting statistical analyses or just data management type of support?" "Will we be providing graphical output such as maps or plots?" "Are we going to be transferring data or accessing a server database?" and "Will we be reading in various file formats like Microsoft Excel or Access?" You may decide that the base system will be enough and you will not have the need to buy additional packages. You should find out what your organization is planning to support with the SAS® programming group and then consult a SAS® sales expert to help guide you in the right direction.

SAS® VALIDATION AND SOPS

After SAS® has been installed in your organization you can technically start programming right away. However, if you are planning to submit data analyses and reports to the FDA there are a few more items that need to be in place. The next step the Biostatistics Branch of USAMMDA took was to develop Standard Operating Procedures (SOPs) regarding SAS® programming. One very important SOP is how your organization plans to validate and update SAS®.

Documenting software validation is important to the FDA because it ensures that the output and reports sent to them and used for final reporting is exactly what was programmed and expected if the same code was used on another validated computer environment. This allows the output to be reproduced with the exact same results. SAS® conveniently provides Installation Qualification (IQ) and Operation Qualification (OQ) scripts that can be used to validate the version of SAS®. However, procedures need to document when, who and how often the IQ and OQ scripts are run, who reviews and stores these log and output files, and where those files will be stored and archived. The procedure should also state who is responsible for checking with the SAS® Institute if an error occurs. Each organization will also want to develop a procedure for the performance qualification or SAS® program validation.

Validating a SAS® program involves requirements and specifications gathering, program testing, and program maintenance (Howard). Adequate documentation must be kept with each program in order to prove to the FDA that the programs have been sufficiently tested and that the output provided to them is an accurate and complete representation of the intended output. Within the USAMMDA Biostatistics Branch, SOPs were developed to document the procedure for each of the steps included in SAS® program validation. Accompanying forms to collect required information were also developed. This provides the programmer with adequate information prior to the start of programming and testing. Each program is given a risk level of low, medium, or high, depending upon the complexity and intended use of the program and output (Benze). Testing is then performed and documented according to the risk associated with it. Once the program has passed testing, the program along with all documentation is archived according to the respective SOP.

Along with SOPs for program validation, we also developed an SOP regarding the handling of study data once we receive a data transfer from a contractor. This SOP specifies who retains the data CD, where the electronic study data is kept, who is notified of the receipt, and who has access to the study data.

We also implemented an SOP regarding the network directory structure and how to store SAS® programs, output, macros, data, and formats. Since we are a small group and have a large work load, it was important that we develop a network structure that maximizes organization and efficiency. A following section will outline in more detail the method we used for setting up our directory structure.

Another important SOP details how the USAMMDA Biostatistics Branch will back-up the data incase of data loss. All data must be archived so that program outputs can be reproduced if the need arises. The FDA may have the need to audit our organization and it is imperative that we be able to reproduce any program outputs sent to the FDA.

Another useful SOP is one detailing programming standards that should be used when creating SAS® programs. The standards should not be too detailed or stringent to discourage various and creative programming styles, but should

be a minimum set of rules to follow when generating SAS® code. Included in this SOP should be documentation regarding naming SAS® files and standards, how and when to comment your programs, what to include in a SAS® header, and various code formatting styles within the SAS® code such as indenting. These standards will make code review easier by other programmers and reduce the time needed to perform these code reviews.

SETTING UP YOUR SAS® PROGRAMMING DIRECTORY

Another important aspect to setting up your SAS® programming group is to set up your programming directory correctly. An efficient directory structure will allow programmers to easily find the latest data and analysis files, find the program and output that they need to work on, find documents related to the study such as the protocol, statistical analysis plan, or annotated case report forms, and applies the correct formats to the data. If the directory is set up correctly, SAS® programs can potentially be used across multiple protocols with only minor changes. This will greatly increase the efficiency of your programming. Using a SAS® autoexec file for each protocol will also increase the group's efficiency and will ensure that the programs are using the correct data, macros, and formats.

An autoexec file (autoexec.sas) is a SAS® program file which is executed prior to each SAS® program that is run in batch mode. One caveat to an autoexec file is that it must be located in the same directory as the SAS® program file. If SAS® is executed interactively, the autoexec file will need to be run interactively as well. Within the SAS® autoexec file you can specify standard macro variables, directory paths, formatting options, library statements, formats, or any other snippet of code that will be used for multiple programs. For example, I define all my libraries in the autoexec file. That way if I need to change the data location, this change will apply to all programs in that directory. This ensures that all the programs use the same data sets. Also in an autoexec file you can specify the search order for finding macros and data formats. We have set up the search order to start with protocol specific macros and formats before finding more general macros and formats. This is a direct reflection to the directory structure.

For the USAMMDA Biostatistics Branch, our programming directory has a top level general programming folder which contains each pharmaceutical product and general macros and formats. Within each pharmaceutical product, there is a folder for each protocol and product specific macros and formats. Finally, each protocol has a specific folder that contains important study documents (e.g. protocol, case report forms, statistical analysis plan, and table shells), formats, macros, data, SAS® programs, and output files. This structure may vary depending on your individual organizations needs. For example, if you are a CRO then you may want to further organize your files according to the sponsor organization.

SAS® MACRO LIBRARY

The USAMMDA Biostatistics Branch is involved with over a dozen protocols in various stages of development, from pre-clinical to study close-out. We have tried to build in efficiency where ever we could. Building a SAS® macro library is an important step in making your programming group efficient at programming tasks. For any tedious tasks that are repeated numerous times, this is the perfect place for a SAS® macro to step in and streamline the process.

Within the Biostatistics Branch, we have developed macros for converting SAS® transport files into SAS® data sets, appending universal fields (i.e. subject identifiers, subject demographics, and study population groups), importing data from other file formats (i.e. Microsoft Excel, Access, or Comma Separated Value), calculating standard variables (i.e. age, study day) generate table and listing reports, and insert title and footnote statements onto output. We also have a Microsoft Excel file that has all program output titles which is read into each SAS® program with a macro. The macro then assigns the correct title statements to the SAS® output. If the SAS® output titles change, this method makes it easy to update output titles and numbers without having to open every SAS® program file.

As mentioned in the previous section, the most general macros are stored in the root programming directory folder. These are macros that can be used across multiple products and protocols such as an age calculation macro, or listing macro, or a SAS® transport file conversion macro. Protocol specific macros are then stored in a macro folder in each protocol folder. These macros can be variations on global macros and can have the same name as the global macro or they can be specific to only that protocol.

The macro library development process is one that takes time to implement. I do not believe that you can develop an effective macro programming library before generating any SAS® programs. As I was writing SAS® programs, I found there were specific functions that would make sense to implement as a macro. The more SAS® programs that I wrote, the more macros I developed. Initially, it took a great deal of effort to write the macros, but they have saved us a lot of time.

CONCLUSION

There are many ways in which an organization can set up their SAS® programming departments. This paper illustrated one way in which we have assisted USAMMDA in establishing their SAS® Programming group incorporating efficient processes, FDA and ICH compliant policies, and good industry standard practices. Ultimately, it comes down to the needs and requirements for each organization as to how each group will be developed and what processes and standards will be implemented for the group. It is important to become familiar with the FDA and ICH

guidances when developing your group. I would also recommend keeping your regulatory affairs experts involved in the process because they should know many of the details regarding the various FDA and ICH guidances.

REFERENCES

Benze, Keith C. *Risk-Based Approach to SAS® Program Validation*. Morrisville, NC: PharmaSUG 2005 Paper FC04.

Howard, Neil. *Beyond Debugging: Program Validation*. Basking Ridge, NJ: SUGI 28 Paper 58-28

CONTACT INFORMATION

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

Devin J. Hunt
Clinical Research Management
1430 Veterans Drive
Fort Detrick, MD 21702
301-452-0813
dhunt81@gmail.com

Nicole C. Close, PhD
Clinical Research Management
1430 Veterans Drive
Fort Detrick, MD 21702
301-524-4104
nclostat@yahoo.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.