1. PURPOSE

This work instruction describes the statistical programming practices at BeiGene including bulk uploading of documents into Veeva Vault for Submissions (VVS).

2. SCOPE

This work instruction applies to all statistical programming performed by the Global Statistics and Data Science (GSDS) group to support BeiGene clinical studies.

3. ABBREVIATIONS AND DEFINITIONS

3.1 **AWS**: BeiGene's internally hosted Amazon Web Service environment.

3.2 **CRO**: Contract Research Organization.

3.3 **EDC**: Electronic Data Capture, A computerized system designed to collect clinical trial data in electronic form as opposed to paper form.

3.4 **FTP**: File Transfer Protocol

3.5 **Global Statistics and Data Science (GSDS)**: This is a group at BeiGene which includes Statistics, Statistical Programming, Data Management, and Data Science and Operational Excellence.

3.6 **JIRA**: Web-based system to manage system access requests and technical issues.

3.7 **Linux File Share Drive**: and operating system, based on UNIX, that runs on many different hardware platforms and whose source code is publicly available.

3.8 **Python**: Python is an interpreted, high-level, general-purpose open source programming utility.

3.9 **SAP**: Statistical Analysis Plan.

3.10 **SAS Macro**: A computer program written using SAS macro language that can be used across multiple projects within GSDS.

3.11 **sFTP**: Secure File Transfer Protocol – method of securely transferring files between BeiGene and external vendors.

3.12 **TFLs**: Tables, Figures, and Listings.

3.13 **Utility**: A computer script or SAS program template serving in several capacities that can be used across multiple projects within GSDS.

3.14 **VVS**: Veeva Vault for Submissions.

4. RESPONSIBILITIES

4.1 **Head of Statistical Programming or designee** is responsible for ensuring that each SAS user completes the required training and complies with the relevant work instructions.

4.2 **Statistical Programming Product Lead (SPPL)** is responsible for approving the access to the study data and the analysis folders.

4.3 **Study Lead Programmer (SLP)** is responsible for maintaining the records of verification process and managing the overall project or clinical study.

4.4 **Statistical Programmer (SP)** is responsible for completing the required training and following the practices outlined in this work instruction. Statistical Programmer is also responsible for requesting the study data and analysis folders access via GSDS Access Request Form in JIRA.

4.5 **SAS System Administrator (IT)** are responsible for providing technical support and user access management. They need to ensure the users have completed the SAS trainings in LMS, prior to granting permission to the SAS environment.

4.6 **Data Engineering and Analytics (DEA)** is a member of the DSOE department responsible for programming data listings and reports in support of data review, cleaning, and reporting of study data.

4.7 **Statistician** is responsible for TFL shells development and maintenance.

5. INSTRUCTION

5.1 System Access Management

5.1.1 Statistical Programmer

5.1.1.1 AWS is a secured virtual environment in which Statistical Programming and Data Engineering and Analytics (DEA) developed and maintained. User MUST complete required training to get the access. Please refer to WI (VV-QDOC-00381 SAS and AWS User Training) for the process of requesting access and setup of SAS system.

5.1.1.2 Please refer to WI (VV-QDOC-14391 GSDS Data Access Request and Access Management), and SOP (VV-QDOC-12052 GSDS Data Access and Control) for steps to apply for AWS file share access through GSDS Access Request Form in JIRA.

5.2 Statistical Programming and Verification of GSDS Deliverables

5.2.1 Statistical Programmer

5.2.1.1 Performs the statistical programming per the programming practices outlined in Appendix 1 and Appendix 2.

5.2.1.2 Performs verification of the programs that produce datasets and outputs per WI (VV-QDOC-00698 Verification of Statistical Programming Deliverables).

5.2.1.3 Documents the programming development and verification in the Template (VV-QDOC-02099 Verification Documentation Form).

5.3 Contract Research Organization (CRO) Management

5.3.1 Study Lead Programmer

5.3.1.1 Reviews CRO process/SOP on verification/validation of datasets and output to ensure CRO processes are adequate and in compliance.

5.3.1.2 Maintains appropriate communication with CRO, who is responsible for generation of any statistical output including SDTM, Analysis Data, Tables, Figures and Listings as specified in the Statistical Analysis Plan (SAP) and in the contract, to ensure project goals, timelines and output quality are met.

5.3.1.3 Reviews relevant documents created by CROs, such as SDTM specifications, analysis data definition document and TFL shells, to make sure BeiGene standards are being followed.

5.3.1.4 Explains BeiGene's analysis data and output requirements to CRO so that quality outputs are delivered when needed.

5.3.1.5 Ensure appropriate review timeframes are built in the project timelines and communicated to the CRO.

5.4 SAS Data and Output Transfer to and from CRO

5.4.1 Study Lead Programmer

5.4.1.1 Ensure that data/output file transfer between CRO and BeiGene is in file(s)/zip files through an FTP or sFTP portal and store in secured folders. Please refer to WI (VV-QDOC-14393 Statistical Programming sFTP Structure and Data Processing).

NOTE: Configuration and manual transfer of CRO data via FTP or sFTP is restricted only to Study Lead Programmer.

5.5 Study Closeout or Milestone Analysis Completion

5.5.1 Study Lead Programmer

5.5.1.1 When a study CSR has been signed-off or when work on milestone analysis has been completed, verify the following:

- Accuracy and completeness of the Template (VV-QDOC-02099 Verification Documentation Form).
- All files are in appropriate locations

5.5.1.2 For Final and Interim analysis, copy all relevant Production and Development macros/utilities used to the tools folder to ensure exact replication of results.

5.5.1.3 Contact IT Service Desk at AskIT@beigene.com to request to lock down the analysis folder by setting the folder to read-only.

5.6 Submission of Documents to VVS

5.6.1 Pre-Requisites:

5.6.1.1 Submission Binders are created per SOP (VV-QDOC-25071 Submission Dossier Binder Management). Study Lead Programmer does not create submission binders.

5.6.1.2 Edit access to the binder is required.

5.6.1.2.1 Study Lead Programmer needs to request edit access from RA team.

5.6.1.2.2 Study Lead Programmer ensures there is edit access to the binder before starting the upload.

5.6.1.3 Study Lead Programmer ensures each document is named per GRO-JA-RIM-0002 BRIM Data and Document File Naming Standards.

5.6.2 Study Lead Programmer uploads, classifies, assigns metadata, shares and initiates review of the documents per WI (VV-QDOC-00188 Upload/Authoring of Documents in Veeva Vault Submissions).

6. REFERENCES

6.1 Controlled Documents

6.1.1 Form (VV-QDOC-00037 SAS Macro and Utility Change Control)

6.1.2 SOP (VV-QDOC-12052 GSDS Data Access and Control)

6.1.3 SOP (VV-QDOC-25071 Submission Dossier Binder Management)

6.1.4 Template (VV-QDOC-02099 Verification Documentation Form)

6.1.5 Template (VV-QDOC-13241 Tables, Figures, Listings (TFLs) Shells)

6.1.6 WI (VV-QDOC-00170 GSDS Macros and Utilities Validation)

6.1.7 WI (VV-QDOC-00188 Upload/Authoring of Documents in Veeva Vault Submissions)

6.1.8 WI (VV-QDOC-00381 SAS and AWS User Training)

6.1.9 WI (VV-QDOC-00698 Verification of Statistical Programming Deliverables)

6.1.10 WI (VV-QDOC-14391 GSDS Data Access Request and Access Management)

6.1.11 WI (VV-QDOC-14393 Statistical Programming sFTP Structure and Data Processing)

6.2 Regulatory References

6.2.1 N/A

6.3 Other References

6.3.1 GRO-JA-RIM-0002 BRIM Data and Document File Naming Standards

7. APPENDICES

Appendix 1: Programming in US AWS

A1. STATISTICAL PROGRAMMING AREA

1.1 The primary file shares for data storage, and programming are:

/usrfiles/bgcrh/bdata
/usrfiles/bgcrh/build
/usrfiles/udata
/usrfiles/unblind
/usrfiles/bgcrh/support
/usrfiles/archive
……

1.2 Non-restricted EDC and other clinical source data are stored at:

/usrfiles/bgcrh/bdata

1.3 Unblinded/restricted EDC and other clinical source data for use by unblinded personnel are stored at:

/usrfiles/udata

1.4 Analyses programs and outputs for clinical study deliverables that have been completed and closed are located at:

/usrfiles/archive

1.5 Macros are located at:

/usrfiles/support/macros/gmvuat

Once validated, they will be moved under

/usrfiles/support/macros/gmv

1.6 Utility programs that used across projects are located at:

/usrfiles/bgcrh/support/utilities

A2. AWS FILE SHARE FOLDER SECURITY

2.1 Please refer to SOP (VV-QDOC-12052 GSDS Data Access and Control).

A3. PROGRAMMING DIRECTORY STRUCTURE

3.1 The SAS server has two main shares (/usrfiles/bgcrh/build and /usrfiles/bgcrh/bdata). Directories are organized to facilitate analysis preservation. The general structure of /usrfiles/bgcrh/build and /usrfiles/bgcrh/bdata directories for a study are as follows:

3.1.1 Data Storage Area

Project → Study → current/snapshot

/usrfiles/bgcrh/bdata/<project>/<study> share has the following standard directories (names are case sensitive) under each study level:

- current
- rd_maint
- <analysis/task >

3.1.2 Analysis/Task level area:

Project → Study → analysis/task

3.1.3 DEA development area:

cp → blinded → <product> → <study>

cp → unblinded → <product> → <study>

3.1.4 /usrfiles/bgcrh/build/<project>/<study>/< analysis> AND /usrfiles/bgcrh/cp/<blinded/ublinded>/<project>/<study> share the following standard directories (names are case sensitive) under each analysis/task level:

- docs
- exports
- import
- dev
- val
- tools

A4. PRIMARY DIRECTORY DETAILS

The standard folder structure under /bgcrh/build and /bgcrh/bdata drives allows for automation of tasks and should not be deleted or changed without approval from the Study Lead Programmer.

DIRECTORY	PURPOSE
/usrfiles/bgcrh/build/<project>/<study>/<analysis/task> folders:	
/docs/protocol	Directory to store study protocol
/docs/crf	Directory to store CRF
/docs/sap	Directory to store SAP, TFL Mockups, TOC and analysis related documentation
/docs/dmdocs	Directory for ALS, DTP, CCG and other documentation related to data management
/docs/adam	Directory for ADaM specifications
/docs/sdtm	Directory for SDTM specifications
/docs/ stdymgt	Directory for Study Management, assignments and programming study team communications
/export	SAS datasets, define packages, output packages etc. To be sent to outside groups such as CROs or other business partners
/import	Files received from clinical research organizations (CROs) and other vendors and business partners
/dev/pgm/biostat	Development Program folder for Statisticians
/dev/pgm/sdtm	Development Directory for SDTM production programs
/dev/pgm/adam	Development Directory for ADaM production programs
/dev/pgm/tlfs	Development Directory for TFL production programs
/dev/pgm/scripts	Development Directory for production automation scripts
/dev/pgm/adhoc	Development Directory for ad-hoc production programs
/dev/crts/adhoc	Development Directory for ad-hoc production datasets
/dev/crts/sdtm	Development Directory for SDTM production datasets
/dev/crts/adam	Development Directory for ADaM production datasets
/dev/crts/figures	Development Directory for production figures output related datasets. These datasets are produced to facilitated validation process
/dev/crts/listings	Development Directory for production listings output related datasets. These datasets are produced to facilitate validation process
/dev/crts/tables	Development Directory for production tables output related datasets. These datasets are produced to facilitate validation process
/dev/reports/figures	Development Directory for production figures output files

DIRECTORY	PURPOSE
/dev/reports/listings	Development Directory for production listings output files
/dev/reports/tables	Development Directory for production tables output files
/dev/reports/adhoc	Development Directory for production adhoc output files
/dev/reports/intext	Development Directory for production in-text output files
/dev/metadata/adhoc	Development Directory to store metadata for production ad-hoc programs
/dev/metadata/adam	Development Directory to store metadata for production ADaM programs
/dev/metadata/sdtm	Development Directory to store metadata for production SDTM programs
/dev/metadata/figures	Development Directory to store metadata for production figures programs
/dev/metadata/listings	Development Directory to store metadata for production listings programs
/dev/metadata/tables	Development Directory to store metadata for production tables programs
/val/pgm/biostat	Validation Program folder for Statisticians
/val/pgm/sdtm	Directory for SDTM validation programs
/val/pgm/adam	Directory for ADaM validation programs
/val/pgm/tlfs	Directory for TFL validation programs
/val/pgm/scripts	Directory for validation automation scripts
/val/pgm/adhoc	Directory for ad-hoc validation programs
/val/crts/adhoc	Directory for ad-hoc validation datasets
/val/crts/sdtm	Directory for SDTM validation datasets
/val /crts/adam	Directory for ADaM validation datasets
/val /crts/figures	Directory for validation figures output related datasets. These datasets are produced to facilitate validation process
/val /crts/listings	Directory for validation listings output related datasets. These datasets are produced to facilitate validation process
/val /crts/tables	Directory for validation tables output related datasets. These datasets are produced to facilitate validation process
/val/reports/figures	Directory for validation figures output files
/val/reports/listings	Directory for validation listings output files
/val/reports/tables	Directory for validation tables output files
/val/reports/adhoc	Directory for validation adhoc output files
/val/reports/intext	Directory for validation in-text output files
/val/metadata/adhoc	Directory to store metadata for validation ad-hoc programs
/val/metadata/adam	Directory to store metadata for validation ADaM programs

DIRECTORY	PURPOSE
/val/metadata/sdtm	Directory to store metadata for validation SDTM programs
/val/metadata/figures	Directory to store metadata for validation figures programs
/val/metadata/listings	Directory to store metadata for validation listings programs
/val/metadata/tables	Directory to store metadata for validation tables programs
/tools	Directory to store initialization file and study specific macros, catalogs etc.
usrfiles/bgcrh/cp/<blinded/unblinded>/<project>/<study> folders:	
/docs/sftmdr	Specifications and other documentation related to safety medical data review
/docs/pp	Specifications and other documentation related to patient profiles
/docs/sptfrvis	Specifications and other documentation related to Spotfire® visualization
/docs/effmdr	Specifications and other documentation related to efficacy medical data review
/docs/adderv	Specifications for derived dataset to support medical review outputs
/export	Packages and output related to data review to be sent to outside groups such as CROs or other business partners
/import	Files related to data review received from clinical research organizations (CROs) and other vendors and business partners
/dev/pgm/adderv	Development of production derived datasets programs to support medical data review
/dev/pgm/pp	Development of production patient profiles programs
/dev/pgm/sftmdr	Development of production safety data review programs
/dev/pgm/sptfrvis	Development of production Spotfire® visualization programs
/dev/pgm/scripts	Development of production automation scripts
/dev/pgm/effmdr	Development of production efficacy data review programs
/dev/reports/sftmdr	Development Directory for production safety data review output files
/dev/reports/pp	Development Directory for production patient profiles output files
/dev/reports/sptfrvis	Development Directory for production Spotfire® output files
/dev/reports/effmdr	Development Directory for production efficacy data review output files
/dev/crts/adderv	Development Directory for production derived datasets
/dev/crts/sptfrvis	Development Directory for storing datasets from previous and current run for comparison purposes used for Spotfire® visualization
/dev/crts/sftmdr	Development Directory for storing datasets from previous and current run for comparison purposes used for safety data review
/dev/crts/effmdr	Development Directory for storing datasets from previous and current run for comparison purposes used for efficacy data review

DIRECTORY	PURPOSE
/dev/crts/pp	Development Directory for storing datasets from previous and current run for comparison purposes used for patient profiles review
/val/pgm/adderv	Directory for validation derived datasets programs to support medical data review
/val/pgm/pp	Directory for validation patient profiles programs
/val/pgm/sftmdr	Directory for validation safety data review programs
/val/pgm/sptfrvis	Directory for validation Spotfire® visualization programs
/val/pgm/scripts	Directory for validation automation scripts
/val/pgm/effmdr	Directory for validation efficacy data review programs
/val/reports/sftmdr	Directory for validation safety data review output files
/val/reports/pp	Directory for validation patient profiles output files
/val/reports/sptfrvis	Directory for validation Spotfire® output files
/val/reports/effmdr	Directory for validation efficacy data review output files
/val/crts/adderv	Directory for validation derived datasets
/val/crts/adderv	Validation Directory for production derived datasets
/val/crts/sptfrvis	Validation Directory for storing datasets from previous and current run for comparison purposes used for Spotfire® visualization
/val/crts/sftmdr	Validation Directory for storing datasets from previous and current run for comparison purposes used for safety data review
/val/crts/effmdr	Validation Directory for storing datasets from previous and current run for comparison purposes used for efficacy data review
/val/crts/pp	Validation Directory for storing datasets from previous and current run for comparison purposes used for patient profiles review
/prod/pgm/adderv	Directory for validated production derived datasets programs to support medical data review
/tools	Directory to store initialization file and study specific macros, catalogs etc.
/usrfiles/unblind/<project>/<study> folders:	
/unblind	Area for unblinded SAS programming work. Have the same directory structure as /usrfiles/bgcrh/build/<project>/<study>
/usrfiles/bgcrh/bdata/<project>/<study> folders:	
/current	Raw data from clinical database electronic data capture (EDC) system, core labs and other vendor data associated with the trial.
/rd_maint	Raw data used for development of SDTM and/or ADaM datasets that updated on defined schedule. These data usually used for maintenance run.
/current/sftp/archive	Archived SFTP data transfers.

DIRECTORY	PURPOSE
/<analysis/task>	Raw data associated with a specific analysis/task
/usrfiles/udata/<project>/<study> folders:	
Unblinded raw data from clinical database electronic data capture (EDC) system, core labs and other vendor data associated with the trial. These data used by unblinded personnel. Folder structure is the same as under /usrdiles/bgcrh/bdata/<project>/<study>	

A5. SAS RAW DATA

5.1 The clinical study raw data from CROs or EDC systems is stored at the /usrfiles/bgcrh/bdata file share. Only Data Engineering and Analytics (DEA) group has write and modify access to the /usrfiles/bgcrh/bdata share to maintain security of the raw data.

5.2 Raw data from EDC systems are generally downloaded via automatic scripts that are set up by Data Engineering and Analytics (DEA) group. The script securely accesses the vendor FTP site, downloads the zip file and extracts the data into the appropriate raw data folder(s).

5.3 Non-EDC data from external vendors are generally downloaded via automatic scripts that are set up by Data Engineering and Analytics (DEA) group. The script securely accesses the vendor FTP site or BeiGene sFTP site, downloads the file(s) and extracts the data into the appropriate raw data folder(s).

A6. SAS PROGRAM NAMING CONVENTIONS

6.1 All SAS program names must be in lower case and could contain only alphanumeric characters and hyphen.

6.2 All SAS programs should follow the same naming conventions: SDTM or ADaM program should have the same name as dataset (domain) it produces (ex. lb.sas, adsl.sas).

6.3 A table SAS program name should have format t <shell description>-<table or selection description>-<subset description>. For example, all tables using 'Summary of Adverse Events by System Organ Class and Preferred Term' shell should start from t-ae (Assuming shells name is T-AE). Program to create table for all AEs leading to treatment discontinuation should have name t-ae-disc and if subset by age is used then program name should be t-ae-disc-lt60. If program produces multiple outputs based on one shell, then name of the program should be t- <shell description>. Similar: a figure program name should have format f-<shell description>-<figure or selection description>-<subset description> and a listing program name should have format l-<shell description>-<listing or selection description>-<subset description>.

SAS Program	Naming Convention
xx.sas	SDTM Domain SAS program
v-xx.sas	Validation Program for SDTM SAS program
adxx.sas	ADaM SAS program
v-adxx.sas	Validation Program for ADaM SAS Program
t-<sas program name>.sas	Table SAS program
v-t<sas program name>.sas	Validation Table SAS program
l-<sas program name>.sas	Listing SAS program
v-l-<sas program name>.sas	Validation Listing SAS program
f-<sas program name>.sas	Figure SAS program
v-f-<sas program name>.sas	Validation Figure SAS program
a-<sas program name>.sas	Adhoc SAS program
v-a-<sas program name>.sas	Validation Adhoc SAS program
pp-<sas program name>.sas	Patient Profile SAS program
m-<sas program name>.sas	Macro SAS Program
Ms-<sas program name>.sas	Study Specific Macro Program
b-<batch description).(extension)	Batch script files

A7. OUTPUT NAMING CONVENTIONS FOR TABLES, FIGURES AND LISTINGS

7.1 All SAS output names must be in lower case and should contain only alphanumeric characters and hyphens.

7.2 Output file names should be displayed in according to the Template (VV-QDOC-13241 Tables, Figures, Listings (TFLs) Shells).

A8. ANALYSIS FOLDERS AND ASSOCIATED SUB-FOLDERS NAMING CONVENTIONS

8.1 All folder names must be in lower case and should contain only alphanumeric characters and underscores.

8.2 A new analysis folder should be created to support Interim Analysis, Final Analysis, Data Monitoring Committee Analysis, Analysis to support Presentations and Abstracts, Investigator Brochures, and Safety Monitoring Committee.

8.3 Folder naming convention: Repeated deliverables more than one time per year should be named as <analysis>_<yyyymm> (ex. dmc_201703, dmc_201706). Annual deliverables should be named as <analysis>_<yyyy> (Ex. ib_2017). Medical affair deliverables should be named as <name of conference>_<yyyymm>. The date yyyymm should indicate date of actual meeting or date of conference etc. For analysis use date of data transfer used (Ex. iacsr_20170930, csr_20180430) and identify final analysis folder as final_ <analysis>_<yyyymmdd> (Ex. final_csr_20170930), where yyyymmdd is date of database lock.

A9. SAS PROGRAM CONVENTIONS

9.1 All SAS programs must have the standard header and call the standard initialization init_global.sas macro to set up access for data and output folders. The minimal standard header is displayed below and should contain study name, program name, programmer first and last name, date of program completed. Other fields may be added at the discretion of the Study Lead Programmer.

```
/*************************************************************************
***  Study Name: <protocol name>
***  Program: <program name>.sas
***  Programmer: <first name last name>
***  Date: DDMONYYYY
***
***  Description: <add description of program purpose- optional>
***
*************************************************************************
***  MODIFICATIONS:
***  Programmer:
***  Date:
***  Reason:
***
*************************************************************************/
%inc '/usrfiles/bgcrh/support/utilities/init/init_global.sas';
```

9.2 Each program should have INCLUDE statement to the init_global.sas file located:

/usrfiles/bgsrh/support/utilities/init folder.

A10. SOME PROGRAMMING SETUP DETAILS:

All LIBNAME statements and PATH references are stored in the local initialization file located in the tools folder. The Study Lead Programmer is responsible for ascertaining the accuracy of these paths. Study Lead Programmer should use template init_local.sas file and update for study specific requirements. The init_local.sas template located at /usrfiles/bgsrh/support/utilities/init.

Each program should have INPUT section, PROCESS section and OUTPUT section. It is good programming convention to use comments to label these sections for future references. Study specific analysis parts, sub-setting logic and code sections that need to be highlighted should also be commented as good program documentation practice.

Example:

```
/*******************************************************
*** INPUT datasets and parameters
*******************************************************/
%let pgmid=t_demog;
proc sort data=db.rd_dm out=dm; by subjid; run;

/*******************************************************
*** PROCESS – derive variables from input dataset
*******************************************************/
data dm_vars;
        set dm;
        <derivation code>;
run;
/*******************************************************
*** OUTPUT – output table in RTF and PDF formats
*******************************************************/
%m_rtfrpt(indata=&pgmid);
```

10.1 Follow the general programming naming conventions and standards to create programs.

10.2 Follow table shell specifications provided by Statistician.

10.3 Use Global macros available.

10.4 Study specific macros or new macros should only be created if the Study Lead Programmer approves.

10.5 Use programs that already exist under same or different project as an example.

10.6 Contact Study Lead Programmer for directions.

A11. PRODUCTION RUN OF SAS PROGRAMS

11.1 During program development phase, SAS programs can be executed either interactively or via batch mode.

11.2 For final production run the program must be run in batch mode so that the log file can be permanently generated and be available in the program directory. All production logs should contain no programming errors.

A12. STANDARD MACROS AND UTILITIES

12.1 The Statistical Programming Marco team maintains standard macros and utilities for use in programs. Unless needed specifically, the Statistical Programmer should use the standard macros as part of BeiGene standard programming process. Statistical Programmer should seek permission from Study Lead Programmer before creating any new macros or modifying an existing macro to a study level macro.

12.2 Production standard macros are validated macros and have gone through a validation process. The Validation process of Macros and Utilities is specified in WI (VV-QDOC-00170 GSDS Macros and Utilities Validation).

12.3 Macros in development stage and study level macros are not considered validated. Development and study level macros should never be used in verification programs by a second statistical programmer. Reports that are created using development and study level macros must be verified through program verification without using these macros.

12.4 Macros in development will go through a validation process following WI (VV-QDOC-00170 GSDS Macros and Utilities Validation). Once a macro has gone through validation process it is moved to the Read-Only Production folders by Programming Macros Lead.

A13. MACROS

13.1 Global Production macros are located at: /usrfiles/bgcrh/support/macros/gmv

13.2 The macros in Production folder are validated and the code cannot be modified without going through a change control process per WI (VV-QDOC-00170 GSDS Macros and Utilities Validation). Global scale macros have prefix 'm_'.

13.3 Development macros are located at: /usrfiles/bgcrh/support/macros/gmvuat

13.4 Global Macros in development folder are not validated and are still undergoing changes.

13.5 Study level macros are located at:
/usrfiles/bgcrh/build/<project>/<study>/<analysis/task>/tools

13.6 Study level macros are not validated. Study level macros have the prefix 'ms_'. Statistical Programmers should confirm with Study Lead Programmer to make any change to study level macros

13.7 Statistical Programmers should not make changes to any macros under /usrfiles/bgcrh/support/macros folder without consent from Head of Statistical Programming.

13.8 Process for Modifications or Creation of Study Level Macro:

- Need for modification or new macro should be communicated to Study Lead Programmer prior to any work being done.
- Macro should be created under the Tools folder.
- Macro should have prefix of ms.
- Macro validation is not needed. Programs using the study level macro should be verified by verification programs that do not use the study level macro.

13.9 Process for Modifications of Development Level Macros:

- Modification needed should be communicated to Study Lead Programmer, Programming Macros Lead and Head of Statistical Programming.
- Head of Statistical Programming will delegate the responsibility for making the appropriate change.
- Development level macros after finalization of code will be validated per WI (VV-QDOC-00170 GSDS Macros and Utilities Validation) and moved to Production level.

13.10 Process for Modifications to Global Production Macros:

- Modification needed should be communicated to Study Lead Programmer, Programming Macros Lead and Head of Statistical Programming.
- Head of Statistical Programming will delegate the responsibility for making the appropriate change. A Validation Team leader and Tester will be appointed by the Programming Macros Lead. The Form (VV-QDOC-00037 SAS Macro and Utility Change Control) will need to be completed. A copy of the macro is created in the macro development folder which is used for the code changes.

- When code is updated, consideration should be made to ensure that change in macro code is backward compatible so that legacy programs can still be executed. If backward compatibility is not possible as decided by the Head of Statistical Programming, then a new macro should be created.
- After the change is made, the macro will be revalidated and moved to the production level in accordance to WI (VV-QDOC-00170 GSDS Macros and Utilities Validation).

13.11 The standard initialization file will look for macros in the following order:

- Macros under Tools folder (<project>/<study>/<analysis/task>/tools)
- Development Macros
- Production Macros
- Legacy Macros

13.12 Standard utilities are located at: /usrfiles/bgcrh/support/utilities

13.13 Utility documentation is located at:

/usrfiles/bgcrh/build/gmpdev/macros/<utility name>/docs

13.14 Statistical Programmers should never make changes to any global level utilities without consent from Head of Statistical Programming.

13.15 Study level utility programs can be modified by statistical programmers after gaining consent from the Study Lead Programmer.

A14. DATA MANAGEMENT AND MEDICAL DATA REVIEW SUPPORT

14.1 Programmers support data review by the Medical Monitors and Data Management groups by providing Listings and Reports specified by the Medical Monitors. Reports generated are not verified by the Programming group. These reports are verified against EDC data by the Medical Monitors or Data Managers for accuracy before being used to make any data related decisions.

14.2 Some of the typical reports generated are:

Report Name	Report Description
Medical Data Review Listings (MDR)	Listings of raw data to help data managers and medical monitors identify data issues and detect data signals.

A15. SOFTWARE(S) AVAILABLE WITHIN THE GSDS GROUP

15.1 SAS Enterprise Guide and SAS Universal Viewer

15.2 Other Software

R Studio, SAS IML, Pinnacle 21, Python etc. are also available within the GSDS department. Please ask your manager for details on what is available and how these can be accessed and utilized.

ExamDiff, Beyond Compare, MobaXterm and Spotfire are available by default to user's AWS WorkSpaces.

Appendix 2: Programming in China AWS

A1. STATISTICAL PROGRAMMING AREA

1.1 The primary file shares for data storage, and programming are:

/usrfiles/bdata
/usrfiles/udata
/usrfiles/sp/build
/usrfiles/sp/prod
/usrfiles/sp/support

1.2 Non-restricted EDC and other clinical source data are stored at:

/usrfiles/bdata

1.3 Unblinded/restricted EDC and other clinical source data for use by unblinded personnel are stored at:

/usrfiles/udata

1.4 Analyses programs and outputs for clinical study deliverables that have been completed and closed are located at:

/usrfiles/sp/prod

1.5 Macros are located at:

/usrfiles/sp/support/macros/gmvuat

Once validated, they will be moved under:

/usrfiles/sp/support/macros/gmv

1.6 Utility programs that used across projects are located at:

/usrfiles/sp/support/utilities

A2. AWS FILE SHARE FOLDER SECURITY

2.1 Please refer to SOP (VV-QDOC-12052 GSDS Data Access and Control).

A3. PROGRAMMING DIRECTORY STRUCTURE

3.1 The SAS server has two main shares (/usrfiles/sp/build and /usrfiles/bdata). Directories are organized to facilitate analysis preservation. The general structure of /usrfiles/sp/build and /usrfiles/ /bdata directories for a study are as follows:

3.1.1 Analysis/Task level area:

Project → Study → analysis/task

/usrfiles/sp/build/<project>/<study>/<analysis/task> share has the following standard directories (names are case sensitive) under each analysis/task level:

- docs
- exports
- import
- dev
- val
- tools

3.1.2 Data Storage Area

3.1.3 Project → Study → current/snapshot

/usrfiles/bdata/<project>/<study> share has the following standard directories (names are case sensitive) under each study level:

- current
- rd_maint
- <analysis/task >

A4. PRIMARY DIRECTORY DETAILS

The standard folder structure under usrfiles/sp/build and usrfiles/bdata drives allows for automation of tasks and should not be deleted or changed without approval from the Study Lead Programmer.

DIRECTORY	PURPOSE
/usrfiles/sp/build/<project>/<study>/<analysis/task> folders:	
/docs/protocol	Directory to store study protocol
/docs/crf	Directory to store CRF

DIRECTORY	PURPOSE
/docs/sap	Directory to store SAP, TFL Mockups, TOC and analysis related documentation
/docs/dmdocs	Directory for ALS, DTP, CCG and other documentation related to data management
/docs/adam	Directory for ADaM specifications
/docs/sdtm	Directory for SDTM specifications
/docs/ stdymgt	Directory for Study Management, assignments and programming study team communications
/export	SAS datasets, define packages, output packages etc. To be sent to outside groups such as CROs or other business partners
/import	Files received from clinical research organizations (CROs) and other vendors and business partners
/dev/pgm/biostat	Development Program folder for Statisticians
/dev/pgm/sdtm	Development Directory for SDTM production programs
/dev/pgm/adam	Development Directory for ADaM production programs
/dev/pgm/tlfs	Development Directory for TFL production programs
/dev/pgm/scripts	Development Directory for production automation scripts
/dev/pgm/adhoc	Development Directory for ad-hoc production programs
/dev/crts/adhoc	Development Directory for ad-hoc production datasets
/dev/crts/sdtm	Development Directory for SDTM production datasets
/dev/crts/adam	Development Directory for ADaM production datasets
/dev/crts/figures	Development Directory for production figures output related datasets. These datasets are produced to facilitated validation process
/dev/crts/listings	Development Directory for production listings output related datasets. These datasets are produced to facilitate validation process
/dev/crts/tables	Development Directory for production tables output related datasets. These datasets are produced to facilitate validation process
/dev/reports/figures	Development Directory for production figures output files
/dev/reports/listings	Development Directory for production listings output files
/dev/reports/tables	Development Directory for production tables output files
/dev/reports/adhoc	Development Directory for production adhoc output files
/dev/reports/intext	Development Directory for production in-text output files
/dev/metadata/adhoc	Development Directory to store metadata for production ad-hoc programs
/dev/metadata/adam	Development Directory to store metadata for production ADaM programs

DIRECTORY	PURPOSE
/dev/metadata/sdtm	Development Directory to store metadata for production SDTM programs
/dev/metadata/figures	Development Directory to store metadata for production figures programs
/dev/metadata/listings	Development Directory to store metadata for production listings programs
/dev/metadata/tables	Development Directory to store metadata for production tables programs
/val/pgm/biostat	Validation Program folder for Statisticians
/val/pgm/sdtm	Directory for SDTM validation programs
/val/pgm/adam	Directory for ADaM validation programs
/val/pgm/tlfs	Directory for TFL validation programs
/val/pgm/scripts	Directory for validation automation scripts
/val/pgm/adhoc	Directory for ad-hoc validation programs
/val/crts/adhoc	Directory for ad-hoc validation datasets
/val/crts/sdtm	Directory for SDTM validation datasets
/val /crts/adam	Directory for ADaM validation datasets
/val /crts/figures	Directory for validation figures output related datasets. These datasets are produced to facilitate validation process
/val /crts/listings	Directory for validation listings output related datasets. These datasets are produced to facilitate validation process
/val /crts/tables	Directory for validation tables output related datasets. These datasets are produced to facilitate validation process
/val/reports/figures	Directory for validation figures output files
/val/reports/listings	Directory for validation listings output files
/val/reports/tables	Directory for validation tables output files
/val/reports/adhoc	Directory for validation adhoc output files
/val/reports/intext	Directory for validation in-text output files
/val/metadata/adhoc	Directory to store metadata for validation ad-hoc programs
/val/metadata/adam	Directory to store metadata for validation ADaM programs
/val/metadata/sdtm	Directory to store metadata for validation SDTM programs
/val/metadata/figures	Directory to store metadata for validation figures programs
/val/metadata/listings	Directory to store metadata for validation listings programs
/val/metadata/tables	Directory to store metadata for validation tables programs
/tools	Directory to store initialization file and study specific macros, catalogs etc.
/usrfiles/bdata/<project>/<study> folders:	

DIRECTORY	PURPOSE
/current	Raw data from clinical database electronic data capture (EDC) system, core labs and other vendor data associated with the trial.
/rd_maint	Raw data used for development of SDTM and/or ADaM datasets that updated on defined schedule. These data usually used for maintenance run.
/current/sftp/archive	Archived SFTP data transfers.
/<analysis/task>	Raw data associated with a specific analysis/task
/usrfiles/udata/<project>/<study> folders:	
Unblinded raw data from clinical database electronic data capture (EDC) system, core labs and other vendor data associated with the trial. These data used by unblinded personnel. Folder structure is the same as under /usrfiles/bdata/<project>/<study>	

A5. SAS RAW DATA

5.1 The clinical study raw data from CROs or EDC systems is stored at the /usrfiles/bdata file share. Only Data Engineering and Analytics (DEA) group has write and modify access to the /usrfiles/bdata share to maintain security of the raw data.

5.2 Raw data from EDC systems are generally downloaded via automatic scripts that are set up by Data Engineering and Analytics (DEA) group. The script securely accesses the vendor FTP site, downloads the zip file and extracts the data into the appropriate raw data folder(s).

5.3 Non-EDC data from external vendors are generally downloaded via automatic scripts that are set up by Data Engineering and Analytics (DEA) group. The script securely accesses the vendor FTP site or BeiGene sFTP site, downloads the file(s) and extracts the data into the appropriate raw data folder(s).

A6. SAS PROGRAM NAMING CONVENTIONS

6.1 All SAS program names must be in lower case and could contain only alphanumeric characters and hyphen.

6.2 All SAS programs should follow the same naming conventions: SDTM or ADaM program should have the same name as dataset (domain) it produces (ex. lb.sas, adsl.sas).

6.3 A table SAS program name should have format t <shell description>-<table or selection description>-<subset description>. For example, all tables using 'Summary of Adverse Events by System Organ Class and Preferred Term' shell should start from t-ae (Assuming shells name is T-AE). Program to create table for all AEs leading to treatment discontinuation should have name t-ae-disc and if subset by age is used then program name should be t-ae-disc-lt60. If program produces multiple outputs based on one shell, then name of the program should be t- <shell description>. Similar: a figure program name should have format f-<shell description>-<figure or selection description>-<subset description> and a listing program name should have format l-<shell description>-<listing or selection description>-<subset description>.

SAS Program	Naming Convention
xx.sas	SDTM Domain SAS program
v-xx.sas	Validation Program for SDTM SAS program
adxx.sas	ADaM SAS program
v-adxx.sas	Validation Program for ADaM SAS Program
t-<sas program name>.sas	Table SAS program
v-t<sas program name>.sas	Validation Table SAS program
l-<sas program name>.sas	Listing SAS program
v-l-<sas program name>.sas	Validation Listing SAS program
f-<sas program name>.sas	Figure SAS program
v-f-<sas program name>.sas	Validation Figure SAS program
a-<sas program name>.sas	Adhoc SAS program
v-a-<sas program name>.sas	Validation Adhoc SAS program
pp-<sas program name>.sas	Patient Profile SAS program
m-<sas program name>.sas	Macro SAS Program
Ms-<sas program name>.sas	Study Specific Macro Program
b-<batch description).(extension)	Batch script files

A7. OUTPUT NAMING CONVENTIONS FOR TABLES, FIGURES AND LISTINGS

7.1 All SAS output names must be in lower case and should contain only alphanumeric characters and hyphens.

7.2 Output file names should be displayed in according to the Template (VV-QDOC-13241 Tables, Figures, Listings (TFLs) Shells).

A8. ANALYSIS FOLDERS AND ASSOCIATED SUB-FOLDERS NAMING CONVENTIONS

8.1 All folder names must be in lower case and should contain only alphanumeric characters and underscores.

8.2 A new analysis folder should be created to support Interim Analysis, Final Analysis, Data Monitoring Committee Analysis, Analysis to support Presentations and Abstracts, Investigator Brochures, and Safety Monitoring Committee.

8.3 Folder naming convention: Repeated deliverables more than one time per year should be named as <analysis>_<yyyymm> (ex. dmc_201703, dmc_201706). Annual deliverables should be named as <analysis>_<yyyy> (Ex. ib_2017). Medical affair deliverables should be named as <name of conference>_<yyyymm>. The date yyyymm should indicate date of actual meeting or date of conference etc. For analysis use date of data transfer used (Ex. iacsr_20170930, csr_20180430) and identify final analysis folder as final_ <analysis>_<yyyymmdd> (Ex. final_csr_20170930), where yyyymmdd is date of database lock.

A9. SAS PROGRAM CONVENTIONS

9.1 For each <project>/<study>/<analysis/task>, the standard initialization program init_global.sas to set up access for data and output folders is stored in the tools folder. When run program in the interactive model, it requires to run this program first after SEG connects to SAS server. When run program in the batch model, this program should be called in the autoexec option.

9.2 All SAS programs must have the standard header. The minimal standard header is displayed below and should contain study name, program name, programmer first and last name, date of program completed. Other fields may be added at the discretion of the Study Lead Programmer.

```
/*************************************************************************
***  Study Name: <protocol name>
***  Program: <program name>.sas
***  Programmer: <first name last name>
***  Date: DDMONYYYY
***
***  Description: <add description of program purpose- optional>
***
*************************************************************************
***  MODIFICATIONS:
***  Programmer:
***  Date:
***  Reason:
***
*************************************************************************/
```

A10. SOME PROGRAMMING SETUP DETAILS:

All LIBNAME statements and PATH references are stored in the local initialization file located in the tools folder. The Study Lead Programmer is responsible for ascertaining the accuracy of these paths. Study Lead Programmer should use template init_local.sas file and update for study specific requirements. The init_local.sas template located at /usrfiles/sp/support/utilities/init.

Each program should have INPUT section, PROCESS section and OUTPUT section. It is good programming convention to use comments to label these sections for future references. Study specific analysis parts, sub-setting logic and code sections that need to be highlighted should also be commented as good program documentation practice.

Example:

```
/*****************************************************
*** INPUT datasets and parameters
*****************************************************/
%let pgmid=t_demog;
proc sort data=db.rd_dm out=dm; by subjid; run;

/*****************************************************
*** PROCESS – derive variables from input dataset
*****************************************************/
data dm_vars;
        set dm;
        <derivation code>;
```

```
run;
/*****************************************************
*** OUTPUT – output table in RTF and PDF formats
*****************************************************/
%m_rtfrpt(indata=&pgmid);
```

10.1 Follow the general programming naming conventions and standards to create programs.

10.2 Follow table shell specifications provided by Statistician.

10.3 Use Global macros available.

10.4 Study specific macros or new macros should only be created if the Study Lead Programmer approves.

10.5 Use programs that already exist under same or different project as an example.

10.6 Contact Study Lead Programmer for directions.

A11. PRODUCTION RUN OF SAS PROGRAMS

11.1 During program development phase, SAS programs can be executed either interactively or via batch mode.

11.2 For final production run the program must be run in batch mode so that the log file can be permanently generated and be available in the program directory. All production logs should contain no programming errors.

A12. STANDARD MACROS AND UTILITIES

12.1 The Statistical Programming Marco team maintains standard macros and utilities for use in programs. Unless needed specifically, the Statistical Programmer should use the standard macros as part of BeiGene standard programming process. Statistical Programmer should seek permission from Study Lead Programmer before creating any new macros or modifying an existing macro to a study level macro.

12.2 Production standard macros are validated macros and have gone through a validation process. The Validation process of Macros and Utilities is specified in WI (VV-QDOC-00170 GSDS Macros and Utilities Validation).

12.3 Macros in development stage and study level macros are not considered validated. Development and study level macros should never be used in verification programs by a second statistical programmer. Reports that are created using development and study level macros must be verified through program verification without using these macros.

12.4 Macros in development will go through a validation process following WI (VV-QDOC-00170 GSDS Macros and Utilities Validation). Once a macro has gone through validation process it is moved to the Read-Only Production folders by Programming Marco Lead.

A13. MACROS

13.1 Global Production macros are located at: /usrfiles/sp/support/macros/gmv

13.2 The macros in Production folder are validated and the code cannot be modified without going through a change control process per WI (VV-QDOC-00170 GSDS Macros and Utilities Validation). Global scale macros have prefix 'm_'.

13.3 Development macros are located at: /usrfiles/sp/support/macros/gmvuat

13.4 Global Macros in development folder are not validated and are still undergoing changes.

13.5 Study level macros are located at: /usrfiles/sp/build/<project>/<study>/<analysis/task>/tools

13.6 Study level macros are not validated. Study level macros have the prefix 'ms_'. Statistical Programmers should confirm with Study Lead Programmer to make any change to study level macros.

13.7 Statistical Programmers should not make changes to any macros under /usrfiles/sp/support/macros folder without consent from Head of Statistical Programming.

13.8 Process for Modifications or Creation of Study Level Macro:

- Need for modification or new macro should be communicated to Study Lead Programmer prior to any work being done.
- Macro should be created under the Tools folder.
- Macro should have prefix of ms_.
- Macro validation is not needed. Programs using the study level macro should be verified by verification programs that do not use the study level macro.

13.9 Process for Modifications of Development Level Macros:

- Modification needed should be communicated to Study Lead Programmer, Programming Marco Lead and Head of Statistical Programming.
- Head of Statistical Programming will delegate the responsibility for making the appropriate change.
- Development level macros after finalization of code will be validated per WI (VV-QDOC-00170 GSDS Macros and Utilities Validation) and moved to Production level.

13.10 Process for Modifications to Global Production Macros:

- Modification needed should be communicated to Study Lead Programmer, Programming Marco Lead and Head of Statistical Programming.
- Head of Statistical Programming will delegate the responsibility for making the appropriate change. A Validation Team leader and Tester will be appointed by the Programming Marco Lead. The Form (VV-QDOC-00037 SAS Macro and Utility Change Control) will need to be completed. A copy of the macro is created in the macro development folder which is used for the code changes.
- When code is updated, consideration should be made to ensure that change in macro code is backward compatible so that legacy programs can still be executed. If backward compatibility is not possible as decided by the Head of Statistical Programming, then a new macro should be created.
- After the change is made, the macro will be revalidated and moved to the production level in accordance to WI (VV-QDOC-00170 GSDS Macros and Utilities Validation).

13.11 The standard initialization file will look for macros in the following order:

- Macros under Tools folder (<project>/<study>/<analysis/task>/tools)
- Development Macros
- Production Macros
- Legacy Macros

13.12 Standard utilities are located at: /usrfiles/sp/support/utilities

13.13 Utility documentation is located at:

/usrfiles/sp/build/gmpdev/macros/<utility name>/docs

13.14 Statistical Programmers should never make changes to any global level utilities without consent from Head of Statistical Programming.

13.15 Study level utility programs can be modified by statistical programmers after gaining consent from the Study Lead Programmer.

A14. SOFTWARE(S) AVAILABLE WITHIN THE GSDS GROUP

14.1 SAS Enterprise Guide and SAS Universal Viewer

14.2 Other Software

R Studio, SAS IML, Pinnacle 21, Python etc. are also available within the GSDS department. Please ask your manager for details on what is available and how these can be accessed and utilized.

ExamDiff, Beyond Compare, MobaXterm and Spotfire are available by default to user's AWS WorkSpaces.

8. DOCUMENT HISTORY PAGE

Version	Effective Date	Brief Description of Change
1.0	18 Aug 2017	New work instruction to describe the statistical programming process and practices.
2.0	13 May 2019	Sections 5.8 and 5.9: Added instructions to request to install Adobe Standard tool on user's AWS WorkSpaces, and explicitly denying a user access to a specific study folder(s) on BGCRH drive respectively. Remove Appendix B by creating separate document, Template *Statistical Programming Standard Program Documentation Template*.
3.0	12 Jun 2019	• Section 5.7.2: Updated instructions for clarification. • Section 5.9: Updated instructions to include additional type of study folder access requests. • Section 5.9.2: Added instructions for system support • Removed instructions in Sections 5.8.2 and 5.9.2, since they are SAS System Administrator's tasks outlined in WI, *SAS Computing Environment Administration*.
4.0	08 Aug 2019	• Section 3: Added definition for LFTP tool • Section 5.10: Added instructions for LFTP tool
5.0	31 Dec 2019	• Section 3, 5.2.1.2: Removed SAS Studio application references • Section 5: Updated ITservicedesk@beigene.com to AskIT@beigene.com • Section 5.7.2.1: Updated to remove delete reference • Section 5.9: Updated BGCRH references with Linux File Share
6.0	20 Mar 2020	• Section 3: Added definition for Python • Section A15.2: Added instructions for python utility.
7.0	24 Jul 2020	Updated Purpose and added Section 5.8 Veeva Submission Instructions. Updated department name and fixed typos throughout the document.
8.0	14 Jan 2022	• Merged to latest WI template; • Section 4: Added Statistical Programming Product Lead, Data Engineering and Analytics (DEA) and Statistician responsibility for consistence throughout the document. • Section 5.6: Updated Submission of Documents to VVS to consistent with latest referred WI; • Separated appendix into appendix 1 and appendix 2 for programming in US AWS and in China AWS. • Removed 'Document Owner' 'SAS Computing Environment ', 'SAS EG' and 'Classification' from section 3 and section 4 since they are no longer mentioned in the document. • Administrative updates to consistent with other control documents including department name, roles name, etc. Updated section 6 to reflect the references within this WI

Version	Effective Date	Brief Description of Change
9.0	22 Dec 2022	Administrative updates: • Replaced Clinical Data Engineering (CDE) name with Data Engineering and Analytics (DEA). • Updated CP folder paths to reflect the new folder structure. • Replaced Template (VV-QDOC-02561 Submission and Document File Conventions) with GRO-JA-RIM-0002 BRIM Data and Document File Naming Standards.

Document Approvals

Approved Date: 20 Dec 2022

Task: SME Approval Verdict: Approve changes & release	Hui1 Liu, (hui1.liu@beigene.com) Document Owner 19-Dec-2022 04:57:16 GMT+0000

Task: QA Approval Verdict: Approve changes & release	Vijaya Shanti Abburu, (vijaya.abburu@beigene.com) QA Approval 20-Dec-2022 14:32:15 GMT+0000