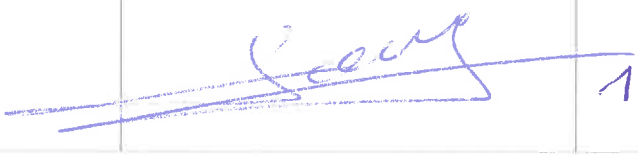
Avenue E. Mounier 83/11
1200 Brussels
Belgium
Tel: +32 2 774 1611
Email: eortc@eortc.be
www.eortc.org

Statistical Programming

ST-003-SOP-01

Version 2.0

ALWAYS REFER TO THE INTRANET TO CHECK THE VALIDITY OF THIS DOCUMENT

Author *Senior Biostatistician* Sandra Collette	Signature: 	Date: *(ex:10-Feb-2017)* 27 Feb 2018
Approved by: *Head Statistics Department* Laurence Collette	Signature: 	Date: *(ex:10-Feb-2017)* 27 Feb 2018
Authorized by: *Director General* Denis Lacombe	Signature: 	Date: *(ex:10-Feb-2017)* 1 MARS 2018

1 PURPOSE

To describe how the statistical analysis programs that generate the EORTC reports are written, validated and documented.

2 SCOPE

The software and programming language that is used for the production of statistical analysis reports or data export/import of EORTC trials is Statistical Analysis System (SAS®, Cary NC, USA) version 9.4 or above. This procedure applies to EORTC staffs who write statistical analysis programs in SAS®.

3 POLICY

All EORTC statisticians and any EORTC staff who develop statistical analysis programs follow the present statistical programming procedure described in this SOP, which results in

- Clarity of the statistical analysis programs
- Maintainability and survivability of the statistical analysis programs
- Ability to transfer analysis programs among statisticians within the EORTC Statistics Department
- Ability for a knowledgeable reader to understand that statistical analysis programs
- Validity and correctness of the output and results produced by the analysis programs

4 DEFINITIONS

- **SAS program**: programs developed using the "base SAS", "SAS Stat" and other SAS tools. When this procedure discusses the verification or validation of a SAS program it is referring to the program code written using SAS and not to the SAS software itself.
- **Data preparation step:** A set of SAS DATA steps, SAS macro calls and SAS PROCEDURES in a SAS program, that manipulates the original trial data as exported from Clinical Database Management System to create the analysis datasets that contain the variables needed for the statistical inference on the study endpoints and for the reporting of the study results in tables or graphics.
- **Data reporting step:** A set of SAS statements (SAS procedures, SAS macro calls…) in a SAS program that performs the statistical inference on the study endpoints and produces the tables, listings and graphics needed for the report of the study results.
- **SAS macro:** Parts of SAS programs that are written in the SAS MACRO language and that enable one to substitute text in a program or to repeatedly perform operations without having to re-write the SAS code.
- **EORTC SAS macro:** validated programming module with a well-defined functionality that can be used across studies, developed in SAS at the EORTC

5 PROCEDURE

5.1 Database export

For each analysis, the locked database including all patients and all data (regardless of date and status) is exported by the Study Statistician in accordance with DM-005-SOP. A single copy of the resulting set of files is stored in the S:\STAT\Trial\ANALYSIS folder on the common drive (see section 6).

5.2 EORTC SAS macros

Standard EORTC SAS macros are developed and then validated by the Statistical Analysts (EORTC staff member who is an expert in SAS programming). The macros are stored on the subdirectory "EORTC macros" on the SAS installation drive on the EORTC server.

The macro validation plan and validation results are documented (ST-003-WIN-01), those documents are reviewed and approved by the Head of the Statistics Department or delegate and filed in protected directory on the EORTC server. Other non-EORTC SAS macros are installed by the Statistical Analysts on a separate subdirectory "extra EORTC macros" on the same server.

5.3 Program Structure

Each SAS analysis program is structured as follows:

- A **header** indicates the study number, the report name and version the program was used for, the program name, the file path of the program, the author's name and the date the program was finalized, as well as the SAS version the program was created with.
- A SAS **libname** declares the permanent library where the calculated dataset will be stored.
- A **data-input step** to import and load the formats and data used for the analysis.
- The **data preparation step** follows, that consists of:
  - Creation of the analysis dataset by selecting all forms with respect to the clinical cut-off date, according to Stat Dept Guideline 7.
  - Creation of analysis-specific formats via proc format
  - Merging of the data including calculation of summary variables for sequential forms using EORTC SAS macros.

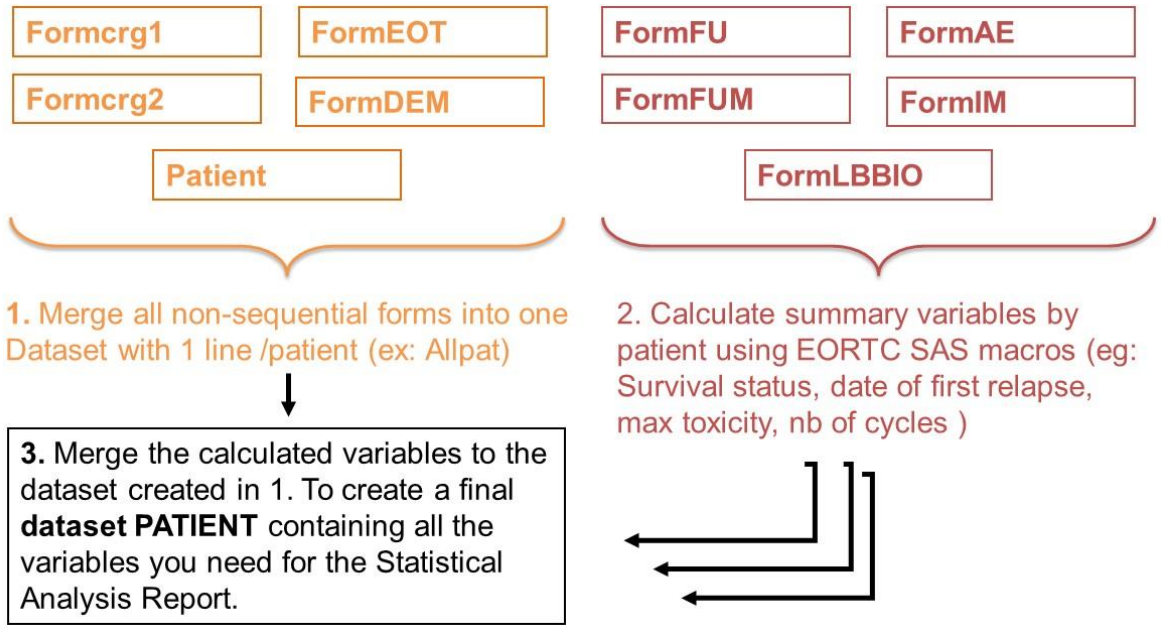


Figure 1: Recommended structure for the data merging step in a SAS program

- Calculations to create flat SAS datasets containing all calculated variables needed for the analysis.

  Calculated SAS datasets, including safety related datasets (clinical adverse events and laboratory data), are stored ***in the permanent library*** pointing to the S:\STAT\Trial\ANALYSIS folder on the common drive. This sequence is represented in Figure 1.

- Optional data checks to verify the consistency and completeness of the data may follow in the next section.

- The **reporting step** consists of a set of commands (mostly SAS macro calls framed within one or several ODS OUTPUT Statements) that produce the core contents of the report: tables and descriptive statistics, survival curves and other statistical inference.

5.4 Assessment that the SAS program/macro runs without errors

Before the reporting step, the following items are checked in the SAS-LOG produced by the SAS program:

- The number of observations in the analysis dataset corresponds to the number of observations in the clinical database.
- There are no remaining ERRORs in the log.
- There are no remaining WARNINGs or "uninitialized variable", or "multiple-multiple merge".

5.5 Program documentation and programming conventions

Programs are documented in such a way that other statisticians or SAS users are able to understand all programming steps and the program's purposes. The documentation is included in the program, in comment sections.

The purpose and content of the program is documented in the program header.

The various steps of the program described in section 5.3 must appear clearly in the program.

An effort is made to ensure that the name, description and purpose of each newly created variable are easy to understand from the program. Comments are used to document all variables for which this is not trivial, i.e. characteristics of patients defined by a categorical variable (unless value labels are self-explanatory), or description of variables defined by combining value of several variables from the original data. Formats and labels are attached to analysis variables as appropriate.

Particular attention is paid to the definition of the patient populations. If analyses are to be performed in subpopulations, it is recommended to create a binary variable representing the inclusion/exclusion from the analysis set. These variables are documented via comments, labels and formats.

It is recommended to use validated EORTC SAS macros when appropriate, especially if macros produce statistical analysis, table, figure or output.

5.6 Version control

Successive reports on the same study are often generated by successive versions of the same program(s).

Whenever a program was used to generate a version of a report that is distributed outside the EORTC Headquarters, the version of the program are kept and a new version created whenever major modifications are required.

When a program is modified, the date (and eventually author) of the modifications are added to the program header. Important modifications of the programming commands are individually documented (change and purpose).

File names of successive versions of a statistical program includes the name of the program and identify the report it has been used for, as well as the version.

5.7 Independent validation of the STAT analysis program

All programs written by Study Statistician to produce interim analysis, final analysis or long term update reports of clinical studies are validated by an Independent Statistician (member of the EORTC Headquarters Statistics Department who is not the Study Statistician, see associated documents).

The Independent Statistician receives the study protocol, the Statistical Analysis Plan if available and a copy of the analysis dataset (1st step of data preparation, see 5.3) and independently writes on his/her own local drive, a SAS program that validates the primary and possibly some secondary endpoints in the population(s) used for the final analysis.

The validation process should start on a test version of the database in order to detect possible data inconsistencies before the database lock. The results of the Study Statistician's program and of the validation program are compared by a one to one comparison of the calculated datasets, and discrepancies are resolved until they reach a consensus point.

In case of subsequent analyses (final analysis after interim analysis or long term update after final analysis), the Independent Statistician re-runs the validation program created for the previous validation and updates it if necessary.

The final approval of the validation by the Independent Statistician has to be done on the locked database, but before the analysis, and both Study Statistician and Independent Statistician report the outcome of the validation on the "Statistical Analysis Program Review and Approval Form" (ST-003-AF-01). The SAS program, the calculated dataset and the resulting report produced by the Independent Statistician are stored by the Study Statistician in the S:\STAT\Trial\ANALYSIS\VALIDATION folder on the common drive. The ST-003-AF-01 form is signed off for approval by the Head or Associate Head of the Statistics Department. An electronic copy (.PDF) of this signed form, with the validation program and the output of the validation program is stored in "SAS" subdirectory of the trial analysis electronic folder.

6 FILING AND ARCHIVING

6.1 S:\STAT\Trial folder

The read/write access to the trial or project folder is restricted to the Study Statistician who works on that particular trial. No read- or write-access is given to a trial folder to the person acting as Independent Statistician for validation purpose, until the validation is complete. A person who has had access to a trial folder can no longer act as Independent Statistician for that study.
The following files should be saved in S:\STAT\Trial\ANALYSIS mm yyyy folder on the common drive:

- Imported original database(s)
- All analysis programs (those SAS programs are run from the S:\STAT\Trial\ANALYSIS mm yyyy folder) and all associated programs (including copies of the validation program to be saved on S:\STAT\Trial\ANALYSIS mm yyyy folder\VALIDATION)
- The calculated datasets used for analysis (including safety related datasets)
- The SAS-LOG of at least the data preparation step is kept with the program output.
- Relevant outputs
- Reports

6.2 Electronic documentation

Only electronic versions of the programs, outputs, reports and associated documentation should be kept.
The minimum documentation to store for each analysis is:

- All versions of the analysis programs,
- The documented validation of the program by the Independent Statistician,
- The calculated datasets used for analysis (including safety related datasets)
- The SAS-LOG of the program data preparation step, and the program output listing (.rtf output);
- All approved official versions of the report resulting from the analyses;
- The graphs used in the reports.

Whenever a study is archived, the S:\STAT\Trial folder is ultimately moved for archiving together with the Trial Master File.

7 RACI MATRIX

Functions / Activities	Statistician	Independent statistician	Statistical Analyst	Head Stat or Associate Head
Statistical Analysis SAS programs				
Database export	A			
Writing SAS programs	A			
SAS analysis program validation	R	R		A
Filling and archiving	A			
SAS macros				
Developing, documenting, validating and maintaining EORTC SAS macros			R	A
Installing and maintaining the non-EORTC SAS macros			R	A
R: responsible; A: accountable; C: consulted; I: informed. Use 'A' when 'A' and 'R' are assigned to the same person.				

8 REFERENCES

None

9 ASSOCIATED DOCUMENTS

Document title	Reference (file name or path)
Statistical Analysis Program Review and Approval Form	ST-003-AF-01
SAS macro validation	ST-003-WIN-01
Clinical cut off	J:\UNIT\Stat\2. Guidelines and Technical documents\ Stat Dept Guideline 7 - Recommendation for clinical cut-off v1.1.pdf
List of independent statisticians	J:\UNIT\Stat\2. Guidelines and Technical documents\Stat Dept Guideline 2 - Path for independent stat v3.0.pdf

10 DOCUMENT HISTORY

Version N°	Brief description of change	Author	Effective date
1.00	Initial release, supersedes WP1304 version 1.1	Laurence Collette	26 Feb 2009
1.1	Update of the process; main changes concerning sections on database, data preparation step and independent validation of the analysis program.	Sandra Collette	13 Mar 2013
1.1	No Change	Sandra Collette	01 Jun 2016
2.0	Addition of RACI matrix No more references to software or electronic tools such as VISTA or GRAPHLIB Clarification about definition of calculated datasets which include safety related dataset	Sandra Collette	01 Mar 2018