

# PhUSE 2007

## Paper AD04

### **MAKS: A validated SAS® Macro Library for Clinical Data in the CDISC SDTM Structure**

Rolf Meinert, Katrin Bhalla-Müller, Jessica Wisniewski, Daniel Wachtlin  
KKS-Mainz, Mainz, Germany ([www.kks-mainz.de](http://www.kks-mainz.de))  
Platform for Medical Research Networks ([www.tmf-ev.de](http://www.tmf-ev.de))

#### **ABSTRACT**

MAKS ("Makros zur Auswertung Klinischer Studien") is a CDISC SDTM based library of validated SAS® macros for reporting data from clinical studies. The macros will be made available to a wide community. Robustness therefore was an important requirement and all macros have only few parameters.

In order to make the SDTM utilizable for the generation of tables and listings it was necessary to amend the SDTM by a so-called variable attribute table (VAT) that contains a set of layout descriptors for SDTM variables. A further file with study-specific layout descriptors like sponsor name and page size can be delivered too.

By end of September 2007, 35 main macros and 80 utility macros have been programmed and validated. Some macros (%AE, %DS, %CM, %LB, %MH, ...) are domain-specific, others (%LBASE, %TBASE, %TIMENUM, %LISTTIME, ...) can display data from several domains.

Conclusion: MAKS is a powerful and validated tool for reporting data from clinical studies using very simple statements. Its only prerequisite is an additional table with variable attributes. The modularity of MAKS will facilitate the addition of new macros in the future.

#### **INTRODUCTION**

The EU GCP-Directive 2001/20/EG states that "*Quality control should be applied to each stage of data handling to ensure that all data are reliable and have been processed correctly.*" This statement also implies the validation of computer programs for the analysis of clinical studies. This requirement means an enormous effort for small companies and organizations, especially for research groups who conduct investigator initiated studies in an academic setting.

According to the ICH Guideline E3 on the structure and contents of clinical study reports a major part of a clinical study report consists of tabulations and listings of are routinely recorded data. These data comprise information on violations of in- and exclusion criteria, demographics, physical examinations, medical history, concomitant medications, administration of study treatment, compliance, vital signs, laboratory data, and adverse events.

In 2003 the Network of Coordination Centers for Clinical Studies (KKS), a network of 12 institutes located at different German university hospitals decided to streamline the effort of SAS programming for clinical studies by the development of a library of validated SAS macros for reporting routine data. This library - named MAKS ("*Makros zur Auswertung Klinischer Studien*") - should be based on a standardized underlying data model. The CDISC SDTM was the only natural candidate for this purpose. Since 2005 the MAKS project has been founded by the Platform for Medical Research Networks (TMF). The TMF comprises about 50 academic German research networks.

#### **PREPARATORY WORK**

##### **USER REQUIREMENTS**

In order to take into account as many of the needs of the involved networks as possible a working group was founded which agreed upon common user requirements and a number of global conventions like producing only outputs in landscape format. The most important requirement was the robustness of the macros because MAKS will be available to a wide community. Therefore, all macros should have as few parameters as possible. For reasons of user friendliness MAKS should not only be able to process 100% SDTM compliant data but also a kind a variation of the SDTM whereby subject characteristics can be stored in domain DM and possible supplemental in the corresponding parent domains. These requirements and conventions were documented in a draft user guide. Finally, the working group collected proposals for standard tables and listings and prioritized these proposals.

##### **TABLE SHELLS**

For each desired representation of clinical data at least one table shell was designed. Later, a minimum and a maximum sample program call was linked to the table shells.

## PhUSE 2007

### FUNCTIONAL SPECIFICATIONS

The draft user guide represented the transfer of the user requirements to functional specifications for the SAS-Programmers. The main macro programs should be designed in a modular fashion to enable the use of small modules - hereafter called "utility programs" - in different main macros. Therefore, a global Program Action Plan (PAP) describing the typical action flow of a main macro was developed. The current version of the global PAP specifies the operation order in which up to 80 utilities are called by a typical main macro. A typical example for a small utility macro is %ISO2SASDATE, a macro that converts a character variable containing dates and times in the ISO8601 format to a SAS datetime variable.

Furthermore, a set of global macro parameters which can be used in all main macros was defined. The only mandatory parameter is &TABNUM, a unique specification of the table number for the study report. All other parameters are optional. The optional parameters comprise &T1, ..., &T5 (titles), &F1, ..., &F5 (footnotes), &POP (analysis population), &PAGEBY (additional variable for a stratification page by page), and &SASCODE (for insertion of free SAS code, e.g. a IF condition). Common macro parameters for table macros are &COLVAR (stratification across columns, default=DM.ARMCD) and &COLBY (further stratification on a 2<sup>nd</sup> level).

### AMENDMENT OF THE SDTM

The main purpose of the CDISC SDTM is the exchange of data between companies and/or organizations and the FDA. The SDTM was not primarily designed as a model for the analysis of data. Hence, it does not provide information about techniques how variables should be analyzed and presented in tables and listings. To overcome this drawback a so-called Global Variable Attribute Table (GLOBVAT) amending the SDTM by additional variable descriptors was designed. The descriptors comprise variable mode (categorical or numerical), variable width (space need in displays), an optional SAS format, labels for headings and values, a sort order for categorical variables and default statistical measures for numeric variables (see Table 1). The GLOBVAT is a permanent SAS-file and it is an essential component of the MAKS library. Users may set up an additional study specific STUDYVAT to modify the default attributes and to proclaim additional previously undefined variables. Furthermore, users have to set up a simple ASCII file called DEFSTUDY.TXT to specify some study specific definitions like for instant a common footnote for all outputs.

**TABLE 1: EXCERPT FROM A VARIABLE ATTRIBUTE TABLE**

vname	domain	vmode	tlabel	tformat	twidth	tstats	orvalue	torder	tvalue
SMOKING	SC	CAT	Smoking Status		12		NONSMOKER	1	Never Smoked
SMOKING	SC	CAT	Smoking Status		12		EXSMOKER	2	Ex-Smoker
SMOKING	SC	CAT	Smoking Status		12		CURRENT SMOKER	3	Current Smoker
BIRTHWGT	SC	NUM	Birth weight (g)	5.1	6	nxsmqr			

### PROGRAMMING AND VALIDATION

#### PROGRAMMING

Before the programming began a Standard Operation Procedure (SOP) for SAS programming was written. Later, another SOP for the validation of SAS programs was added. The SAS programmers had to comply with the detailed programming guideline of the KKS-Mainz in order to avoid heterogeneous programming styles.

The actual programming work started in 2005. At weekly meetings experiences were exchanged, problems were discussed and resolved, and the next steps were planned. The main programming was performed by 2 programmers who dedicated about 50% of their working hours to the MAKS project. This accounts for about 3 person years.

#### VALIDATION

Programmers were not allowed to validate their own programs.

First, the following catalogue of possible validation activities was defined:

- A Reconciliation of input file and paper output for a random sample
- B Review or reproduction of results by a biostatistician
- C Code review (19 standard checks) with respect to the programming guideline
- L Log-file review (10 standard checks)
- O Output review considering specifications (table shells) and overall consistency
- P Reconciliation of the PAP and the actual program flow
- R Reproduction by a 2<sup>nd</sup> programmer for at least 1 example
- T Testing of macros according to a pre-specified formal test plan

Next, it was decided which validation activities had to be applied to which macros. All utility programs went through activities C and L, for more complex programs R and/or T were also applied, and those macros that generate p-values or confidence intervals had to be reviewed by a statistician. Main macros were also validated by means of activities O and P and for listings additionally activity A was performed. A bottom up approach was used: First, all utility programs that do not call another

## PhUSE 2007

program were validated ("level 0" programs). On level 1 programs that only call level 0 utility programs were validated. So far, level 6 is the highest level that has been achieved. All results of the validation activities have been documented in the validation report.

### RESULTS

By end of September 2007, 35 main macros and 80 utility macros comprising about 30,000 lines of code have been programmed and validated. Specific macros are available for the SDTM domains DM & SC, CM, EX, AE, DS, MH, IE, LB, and PE. Some other macros are flexible and can display data from any domain having the findings structure.

So far, the yield of the MAKS project comprises

- 35 validated SAS macros that process the CDISC SDTM,
- a catalogue of 70 examples of program calls and corresponding outputs,
- the documentation,
- the validation plan and the validation report,
- 2 SOPs for SAS-programming for the validation of SAS programs, and
- training courses for users.

### CONCLUSION AND PROSPECTS

The MAKS library offers several advantages to organizations and companies that conduct low budget clinical studies and yet must comply with GCP requirements. If data are provided in the SDTM structure MAKS can produce many desired tables and listings with very little effort by the user. For example, the simple statement

```
%AE ( tabnum=51.1 )
```

will produce a summary table of adverse events stratified for treatment group and ordered by MedDRA system organ class and preferred term. Hence, with MAKS less resources for programming and validation are needed and clinical study reports can be produced much faster.

The catalogue of example outputs is not only a helpful document for novice programmers but also useful in the planning phase of a study because investigators can see in advance how the final tables and listings will look like and choose those representations that they prefer.

Finally, the programming guideline of the KKS-Mainz has proved to be an extraordinary useful tool that noticeably facilitated the comprehension and validation of programs written by other programmers.

The next planned enhancements - along with some minor refinements - are the addition of more flexible macros and the availability of choosing the German language as a standard for all displays.

### REFERENCES

CDISC Study Data Tabulation Model Implementation Guide V3.1.2 Draft

CDISC Study Data Tabulation Model Version 1.2 Draft

both documents can be downloaded from <http://www.cdisc.org>

### CONTACT INFORMATION (HEADER 1)

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

Dr. Rolf Meinert  
KKS Mainz  
Langenbeckstr. 2  
D-55131 Mainz  
Work Phone: (+49) 6131 39 34 637  
Fax: (+49) 6131 39 34 633  
Email: [meinert@kks-mainz.de](mailto:meinert@kks-mainz.de)  
Web: [www.kks-mainz.de](http://www.kks-mainz.de)

Brand and product names are trademarks of their respective companies.