PhUSE 2008

Paper AD07

Using SAS Enterprise Guide Add-In to Enable Guided Statistics

Bas van Bakel, OCS Consulting, Rosmalen, the Netherlands

ABSTRACT

This paper will describe the execution of a project to develop an Add-in for SAS® Enterprise Guide enabling guided statistics. The application is in use at a leading research facility for pre-clinical studies in the Netherlands, NOTOX B.V.

The Add-In loads results from a measuring device and guides the user in the process of determining normality and homogeneity of the available parameters. After this has been determined the Add-In will perform the appropriate statistical procedures and will present the results of these statistical procedures, together with descriptive statistics, in a report. When the Add-In is closed, the intermediate data will still be available within SAS Enterprise Guide and the options within SAS Enterprise Guide can be exploited to perform further (exploratory) analysis on this data. Next to discussing the business drivers, the project approach and the technology used, the Add-in itself will be demonstrated.

INTRODUCTION

This paper consists of the following parts:

- NOTOX B.V.
- Business Drivers
- Project Approach
- Add-In for Lung Function studies
- Technology used
- Project Metrics
- Conclusion

NOTOX B.V.

NOTOX was founded in 1983 and evolved as a privately held company until 2005. In this year NOTOX became part of a worldwide group of autonomous companies providing toxicology research services. This Trans-Atlantic presence was built with a vision to seamlessly address the service needs of international industry partners and ranks in the top 7 CRO's.

NOTOX specializes in personalized contract research and expert consultancy for the registration of pharmaceutical products and (agro)chemicals worldwide. More than 280 well-trained and dedicated specialists and 11,700 m2 of modern purpose built laboratories and offices are available to perform regulatory toxicology studies. Their operations have been endorsed by the GLP Monitoring authorities (OECD/EPA/FDA/JMAFF certification). In addition to the headquarters in 's-Hertogenbosch, the Netherlands, NOTOX has support offices in Japan, Switzerland and the USA.

For more information please visit www.notox.nl

BUSINESS DRIVERS

NOTOX currently owns a validated system which is capable of reporting on the majority of their research data. If the research data is collected in a non-standard way or specific statistical analyses are to be performed on the research data; the validated system does not suffice and reports are generated outside of this validated structure. This requires not only additional statistical programming, but the programs created also need to be validated thoroughly before the results can be sent to clients.

Within each type of study (e.g. Telemetry studies) for which the validated structure does not suffice the collected data is structured similarly. This is because most of the time the same measuring device is used. The statistical analysis within each type of study is also very similar. For example: the first step when analyzing studies investigating the lung function of animals will be to determine the normality and homogeneity of the research data. Depending on the results of the normality and homogeneity testing specific statistical analyses are performed and reports are generated.

Since for each type of study the data is structured similarly and similar statistical analyses are to be performed it is possible to (partially) automate the process of importing and analyzing the data which prevents creating and validating programs for each individual study.
On the other hand, before the exact results of a study are available it is not possible to foresee all exploratory and/or statistical analyses needed and it would be desirable to make it possible for the analyst to perform additional analyses on the imported data which may not be anticipated beforehand.
To fulfil these conditions OCS Consulting has created an Add-In which is an application in which user interface screens guide the user through the partially automated and validated process of importing and analysing the data. This Add-In is to be imported into a SAS Enterprise Guide project which will facilitate additional exploratory analyses.

PROJECT APPROACH

Since the research data for each type of study (Lung Function, Telemetry, etc.) differs and each type of study requires a different kind of analysis, it has been determined to create separate Add-In applications for each type of study.
Before the Add-In application is built meetings are planned with a study director of the particular type of study to discuss in which way the study is to be analysed and which statistical methods are to be used. After this has been determined it is possible to create a study specific 'decision tree' and the lay-out of the user interface screens can be determined. Subsequently the Add-In itself can be created.

ADD-IN FOR LUNG FUNCTION STUDIES

As an example of how an Add-In functions; the developed Add-In for Lung Function studies will be described in this section. First the requested 'decision tree' will be described after which the accompanying Add-In for Lung Function studies will be demonstrated.

ANALYSIS OF LUNG FUNCTION STUDIES

As described in the section project approach, before creating the actual Add-In application the desired statistical methods are determined.
Data of Lung Function studies are stored in a .txt-file which is created by a measuring device. In each Lung Function study several parameters are measured on specific time points for several animals. These animals are included in a (placebo) control group or in one of the active dosing groups.
(Part of) the .txt file of a Lung Function study may therefore look like the file shown in figure 1.

Group	Subject	Range	Cinetic time	Ti	Te	PIF	PEF	TV	EV	RT	MV	f	EIP	EEP	Penh
123456 gr 1	01	pre-dose	--	177	166	28.28	26.40	3.46	3.37	104	606	177	2	16	0.56
123456 gr 1	01	post-dose	+30mn	168	160	31.82	28.52	3.70	3.62	102	684	186	2	15	0.51
123456 gr 1	01	post-dose	+01h	161	154	31.19	28.82	3.56	3.48	98	685	193	2	14	0.53
123456 gr 1	01	post-dose	+01:30h	162	161	31.50	29.59	3.79	3.70	101	711	188	2	15	0.55
123456 gr 1	01	post-dose	+02h	156	150	32.87	32.72	3.83	3.78	95	755	199	2	15	0.59
123456 gr 1	01	post-dose	+02:30h	158	158	28.37	25.61	3.19	3.12	102	611	192	2	15	0.51
123456 gr 1	01	post-dose	+03h	162	154	32.80	31.58	3.81	3.73	97	727	192	2	15	0.57
123456 gr 1	01	post-dose	+03:30h	160	156	26.13	22.81	2.92	2.86	101	560	192	2	14	0.48
123456 gr 1	01	post-dose	+04h	164	157	29.11	27.52	3.47	3.40	101	655	189	2	15	0.54
123456 gr 1	01	post-dose	+04:30h	160	153	29.75	28.43	3.49	3.47	99	671	193	2	14	0.53
123456 gr 1	01	post-dose	+05h	158	150	32.42	32.56	3.90	3.83	95	763	196	2	15	0.59
123456 gr 1	01	post-dose	+05:30h	158	149	31.40	31.85	3.79	3.73	94	746	197	2	15	0.60
123456 gr 1	01	post-dose	+06h	162	153	31.83	33.69	3.92	3.90	96	751	193	2	15	0.63
123456 gr 1	02	pre-dose	--	161	201	17.89	18.61	2.23	2.19	115	372	170	2	20	0.78
123456 gr 1	02	post-dose	+30mn	163	193	17.82	16.97	2.19	2.15	115	369	170	2	19	0.65
123456 gr 1	02	post-dose	+01h	169	205	19.25	18.93	2.45	2.40	119	395	162	2	21	0.72
123456 gr 1	02	post-dose	+01:30h	170	203	18.11	17.14	2.35	2.31	122	382	163	2	20	0.64
123456 gr 1	02	post-dose	+02h	172	219	16.41	15.46	2.15	2.13	129	330	154	2	22	0.66
123456 gr 1	02	post-dose	+02:30h	175	214	17.13	16.90	2.28	2.25	125	353	157	2	22	0.71
123456 gr 1	02	post-dose	+03h	177	210	16.16	15.57	2.19	2.15	125	342	158	2	22	0.66
123456 gr 1	02	post-dose	+03:30h	183	212	16.49	15.55	2.29	2.26	128	353	154	2	21	0.62
123456 gr 1	02	post-dose	+04h	178	204	16.44	15.69	2.25	2.20	123	356	159	2	21	0.63
123456 gr 1	02	post-dose	+04:30h	185	219	15.16	14.26	2.13	2.14	134	317	149	2	22	0.60
123456 gr 1	02	post-dose	+05h	182	211	16.09	14.98	2.26	2.21	128	346	153	2	21	0.61
123456 gr 1	02	post-dose	+05:30h	180	204	15.80	14.79	2.13	2.09	123	335	158	2	21	0.62
123456 gr 1	02	post-dose	+06h	179	202	15.45	15.30	2.11	2.12	123	332	159	2	20	0.64
123456 gr 2	03	pre-dose	--	245	272	15.19	17.48	3.12	3.05	175	364	117	2	24	0.64
123456 gr 2	03	post-dose	+30mn	218	218	14.86	20.60	2.45	2.38	144	338	139	2	19	0.71
123456 gr 2	03	post-dose	+01h	215	210	15.11	19.45	2.30	2.24	142	326	142	3	18	0.63
123456 gr 2	03	post-dose	+01:30h	234	209	14.30	18.76	2.41	2.35	130	329	137	3	21	0.81
123456 gr 2	03	post-dose	+02h	235	208	14.69	19.25	2.47	2.42	129	336	136	3	21	0.82
123456 gr 2	03	post-dose	+02:30h	212	188	16.23	21.66	2.39	2.36	124	359	151	3	18	0.69
123456 gr 2	03	post-dose	+03h	236	200	15.82	20.46	2.60	2.53	127	359	138	3	20	0.75
123456 gr 2	03	post-dose	+03:30h	244	201	16.40	19.41	2.74	2.68	122	372	136	3	21	0.78
123456 gr 2	03	post-dose	+04h	226	187	18.19	20.44	2.74	2.68	118	399	146	3	19	0.67
123456 gr 2	03	post-dose	+04:30h	244	199	17.11	19.66	2.86	2.86	126	390	137	3	20	0.67
123456 gr 2	03	post-dose	+05h	245	204	17.18	19.21	2.99	2.92	126	402	135	2	21	0.69
123456 gr 2	03	post-dose	+05:30h	218	188	18.71	20.86	2.81	2.76	120	418	149	2	19	0.64
123456 gr 2	03	post-dose	+06h	232	197	17.87	20.69	2.97	2.97	126	417	141	2	19	0.66

Figure 1

Each parameter is to be analysed separately. First the normality and homogeneity of the pre-dose values of the parameter are to be determined. If normality and homogeneity are not concluded the user should be able to transform the data according to a pre-specified transformation and/or should be able to exclude outliers, after which the user can determine normality and homogeneity again.
Depending on the results of the normality and homogeneity testing parametric or non-parametric tests should be used for further analyses. Further analyses consist of determining per time point whether the results of each of the active groups differ significantly compared to the control group and (optionally) for each group whether the post-dose values differ significantly from the pre-dose values. If all parameters are analysed the (intermediate) results should be stored in a report.
A (simplified) overview of this 'decision-tree' is shown in figure 2:

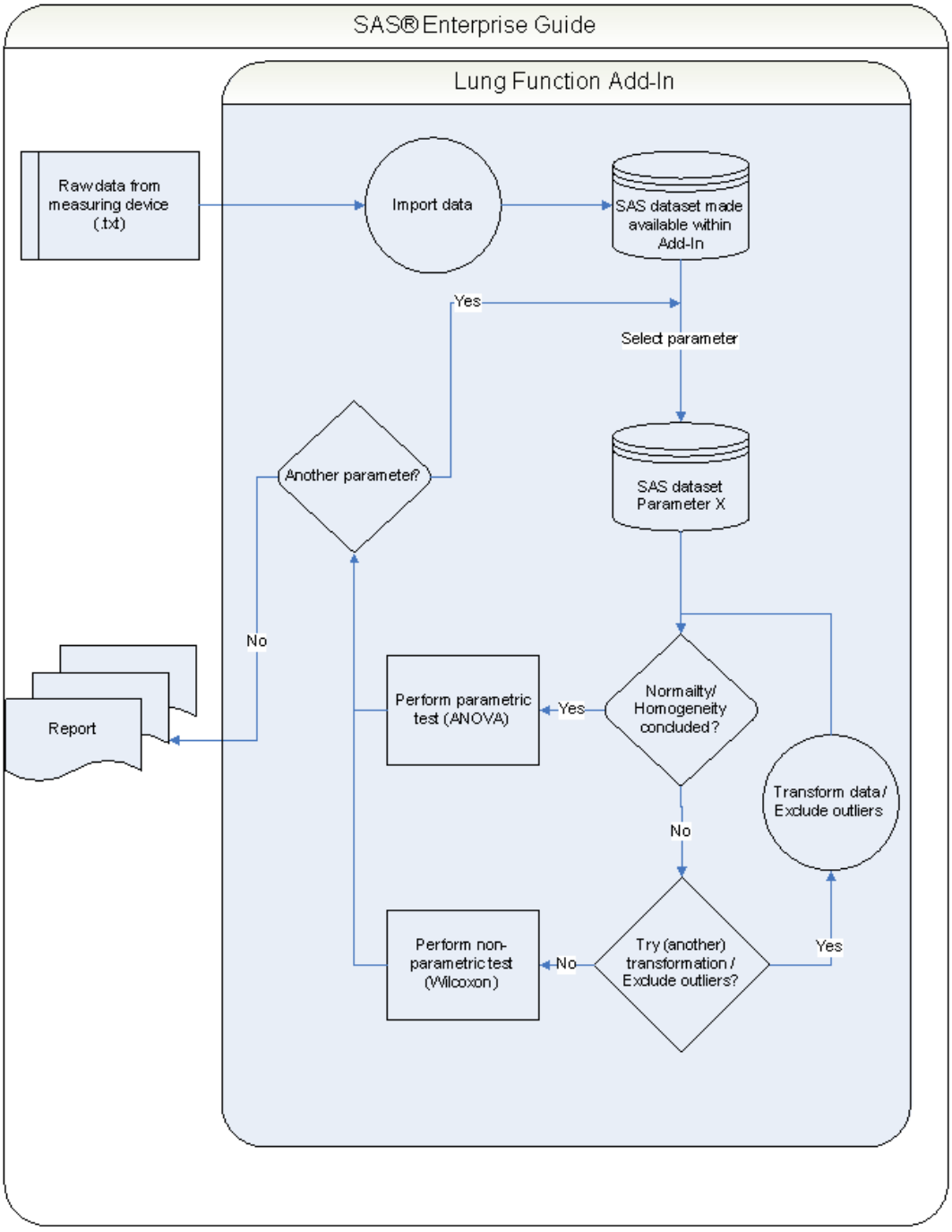

Figure 2

THE ADD-IN AT WORK

A SAS Enterprise Guide Add-In is created to facilitate the steps shown in figure 2. This Add-In (named “SAI-LungFunction”) is made available in the Add-In menu of SAS Enterprise Guide with the help of the Add-In Manager and, when the Add-In application is selected, it will show a browse screen in which the user can specify the .txt file containing the data to be analysed (see figure 3).

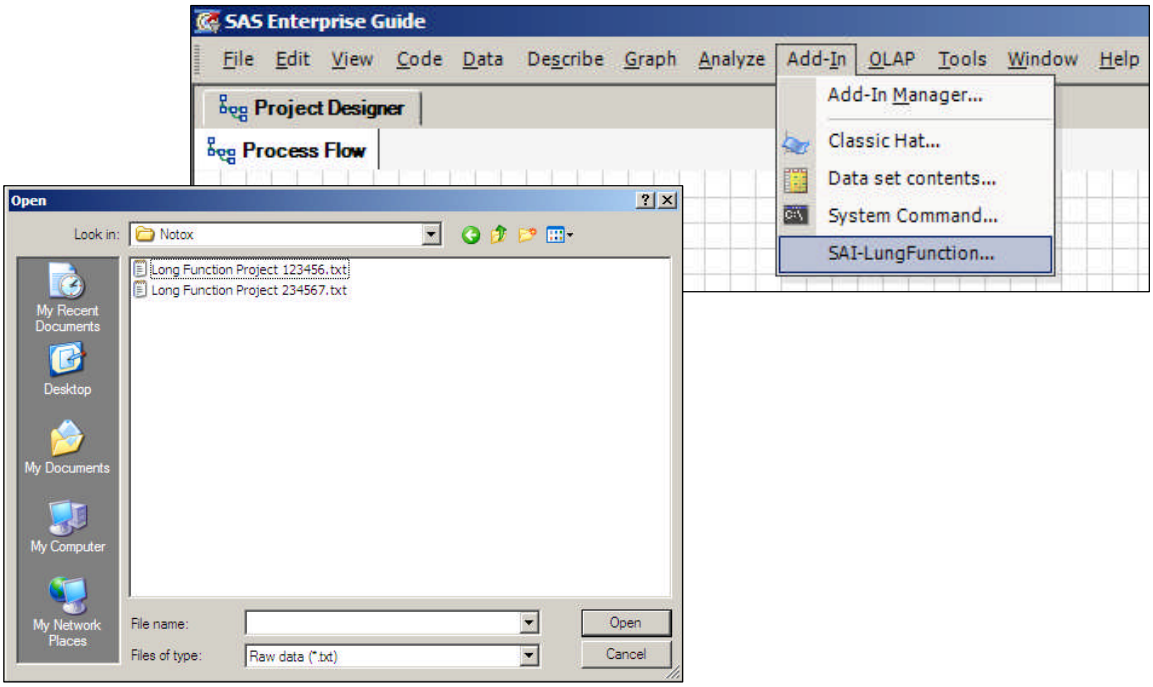

Figure 3

When the input file has been selected the Add-In will import the data and will determine which parameters are available after which it will, by default, perform normality and homogeneity tests for each parameter encountered. The imported data, together with the results of the initial normality and homogeneity tests will be shown in the main user interface screen as shown in figure 4.

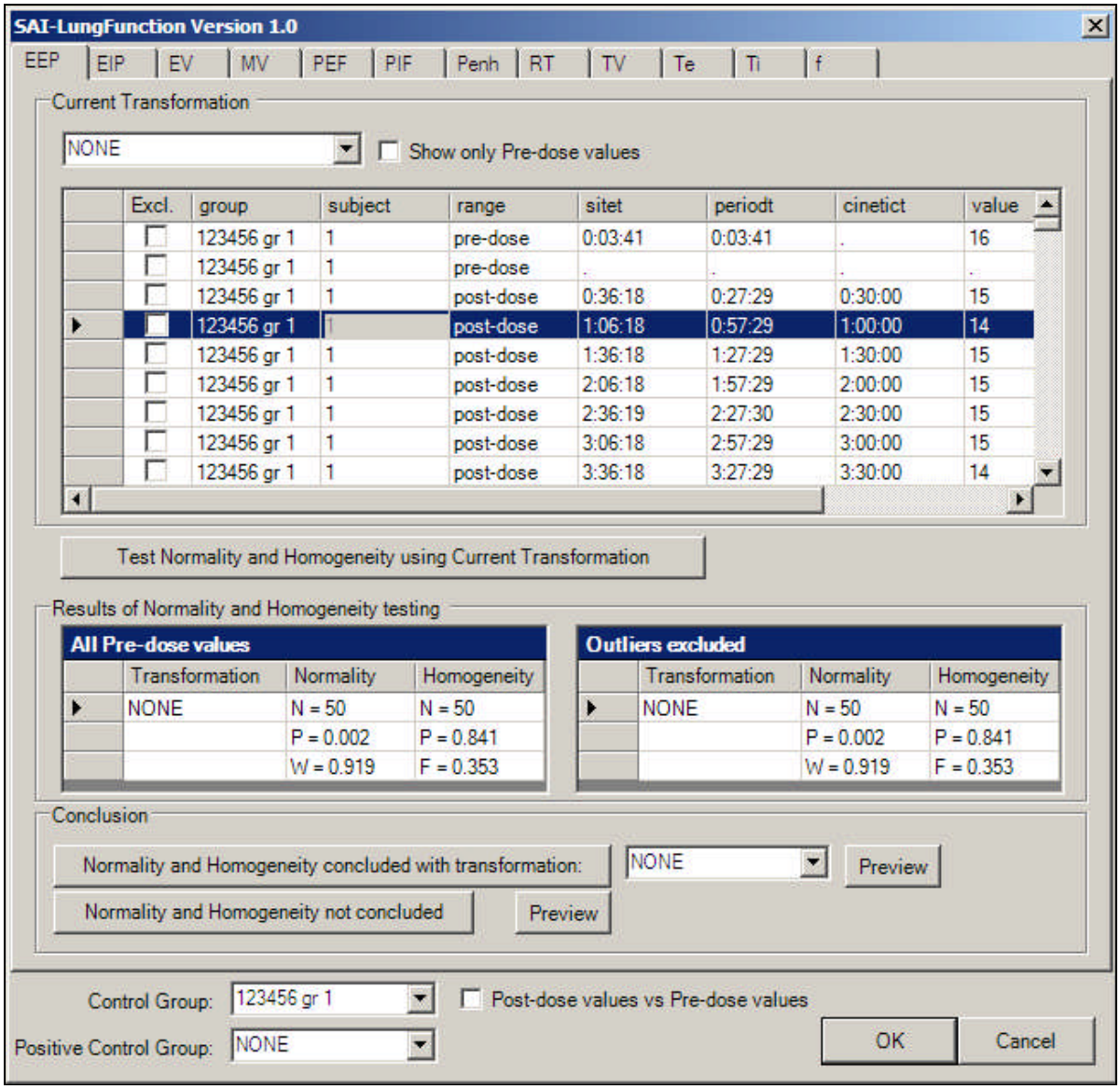

Figure 4

As shown in figure 4 for each parameter encountered ('EEP', 'EIP', 'EV', etc.) a form is created. The imported data of the specific parameter is shown in the data grid at the top of the screen and the results of the normality and homogeneity tests on pre-dose values (the Shapiro Wilk test and Levene (mean of median) test, respectively) are shown in the data grids below.
If normality and homogeneity are not concluded the user has the possibility to select a transformation and/or exclude outliers and test normality and homogeneity again. The user can repeat this step for other transformations and exclusions until he/she concludes normality and homogeneity or until all possible transformations are tested.

The results of the newly performed normality and homogeneity tests will be appended to the previous results as is shown in figure 5. Results of normality and homogeneity tests will be shown for all pre-dose values (data grid on the bottom left) and for the non-excluded pre-dose values (data grid on the bottom right). Transformed values will be appended to the data grid at the top of the form.

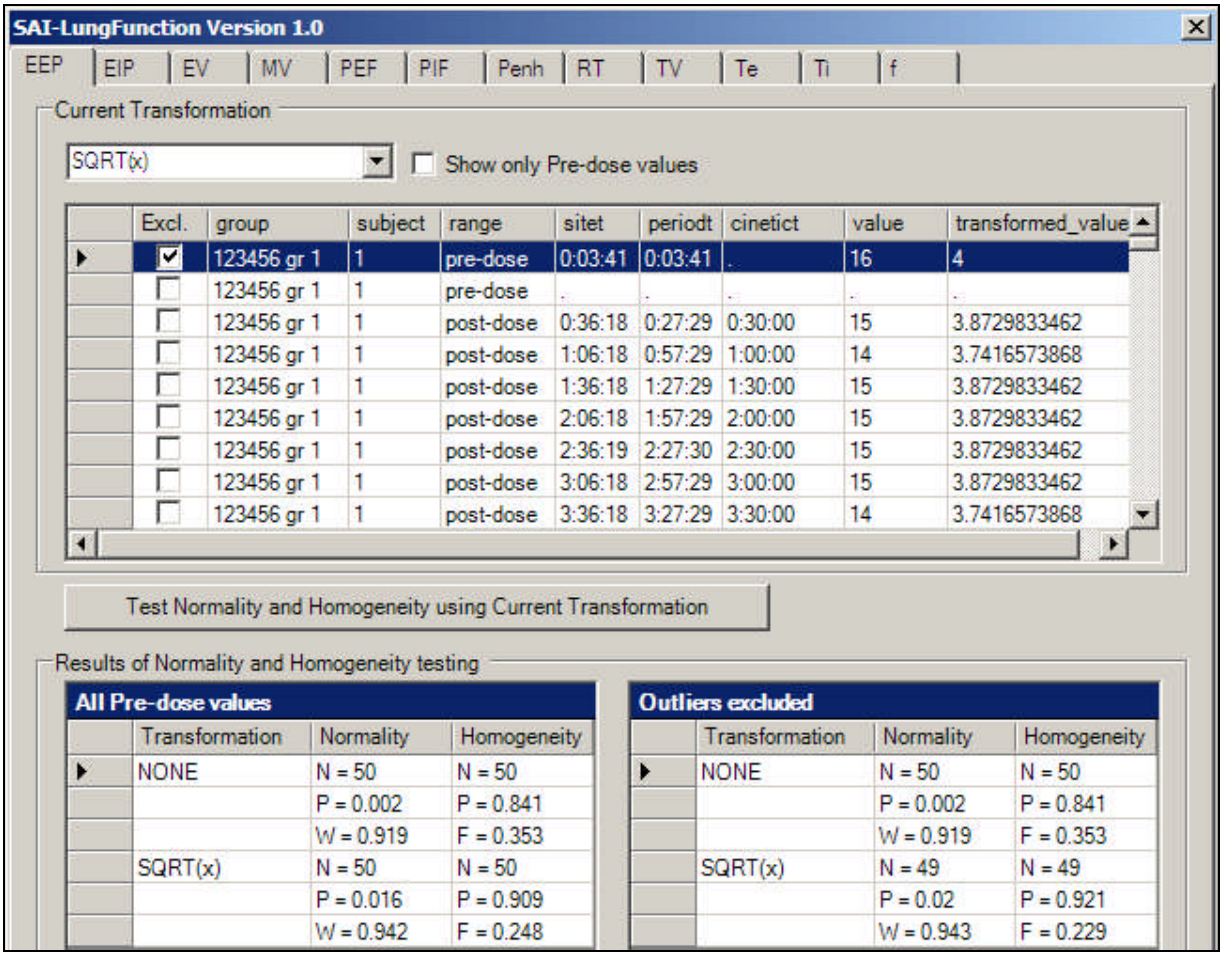

Figure 5

When normality and homogeneity is analysed the user should select either 'Normality and Homogeneity concluded with transformation:' or 'Normality and Homogeneity not concluded' to initiate parametric or non-parametric tests which determine per time point whether the results of each of the active groups differ significantly from the control group and (optionally) for each group whether the post-dose values differ significantly from the pre-dose values.

The user should specify the name of the control group, the positive control group (if available), and whether or not the post-dose values should be compared to the pre-dose values. If normality and homogeneity is concluded and therefore a parametric test is to be performed the transformation used (if applicable) should also be specified (see figure 6).

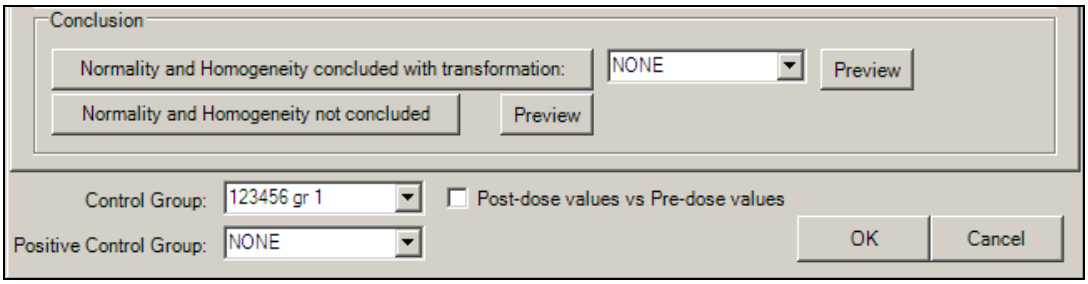

Figure 6

When all required parameters are analysed the user can press the OK button to open a report interface screen in which the user can specify which parameters are to be reported. The report will contain the results of normality and homogeneity tests, the results of the (non-)parametric test comparing the groups/time points and optionally a data listing of the imported data. Figure 7 shows the bookmarked .pdf file with the results of the normality and homogeneity tests and Figure 8 shows an example of a summary table of a specific parameter.

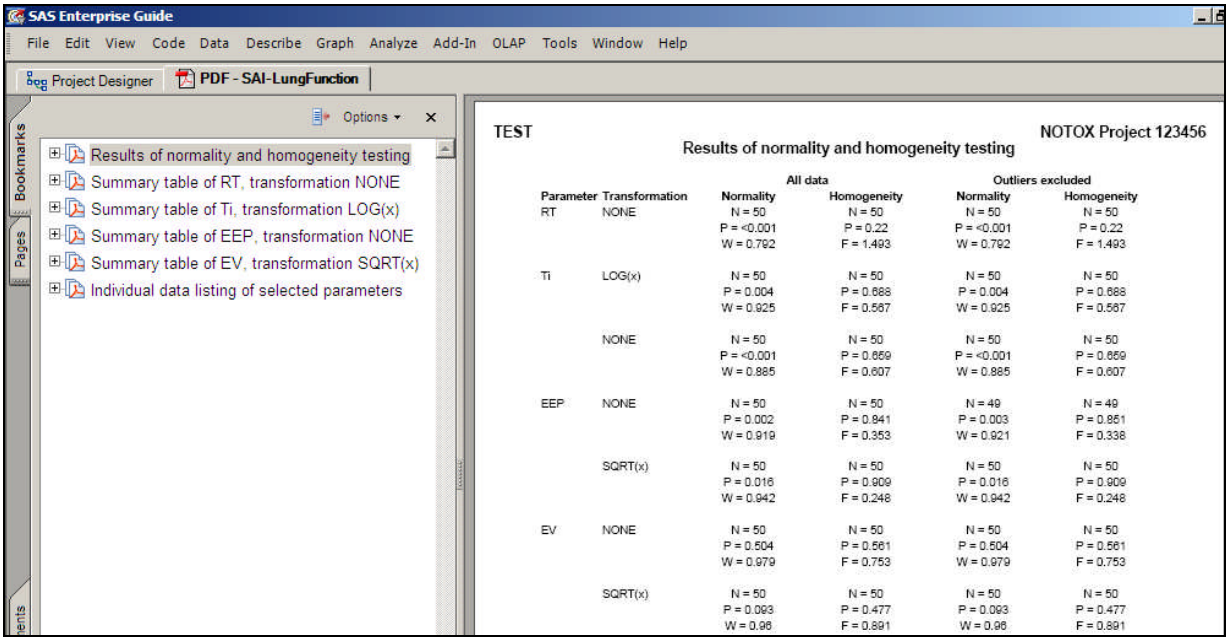

Figure 7

NOTOX Project

Parameter: EEP

		123456 gr 1	123456 gr 2	123456 gr 3	123456 gr 4	123456 gr 5
PREDOSE	MEAN	19.3	18.6	17.8	18.1	16.6
	STD	3.43	3.27	2.90	2.88	2.12
	N	10	10	10	10	10
30 MIN	MEAN	17.9	19.1	17.9	17.2	37.6 #
	STD	1.37	3.11	1.66	1.62	43.9
	N	10	10	10	10	10
1:00 HRS	MEAN	18.9	20.2	17.6	18.7	27.7 *
	STD	2.92	4.29	1.43	2.45	8.77
	N	10	10	10	10	9
1:30 HRS	MEAN	19.7	20.9	18.1	19.0	29.9 *
	STD	3.59	5.84	1.45	2.16	10.4
	N	10	10	10	10	9

*: Significant difference with control group (123456 gr 1) with ANOVA Dunnett test at 5% significance level
#: Significant difference with pre-dose value with ANOVA Dunnett test at 5% significance level

- Page 2 -

15JUL08 14h59

Figure 8

After the report has been generated the Add-In will close and the results will be shown in the Project flow as shown in figure 9. The Add-In will inherit options specified within SAS Enterprise Guide. The style of the reports, for example, will be equal to the style specified in the options of SAS Enterprise Guide. Moreover the report formats that are generated (.html / .rtf / .pdf) are the ones specified in the result output format section.
Datasets created intermediately will be available within the SAS Enterprise Guide session after the Add-In has been closed and these can be used to perform further (exploratory) analyses (see figure 10).
Log files of the SAS code executed will be available.

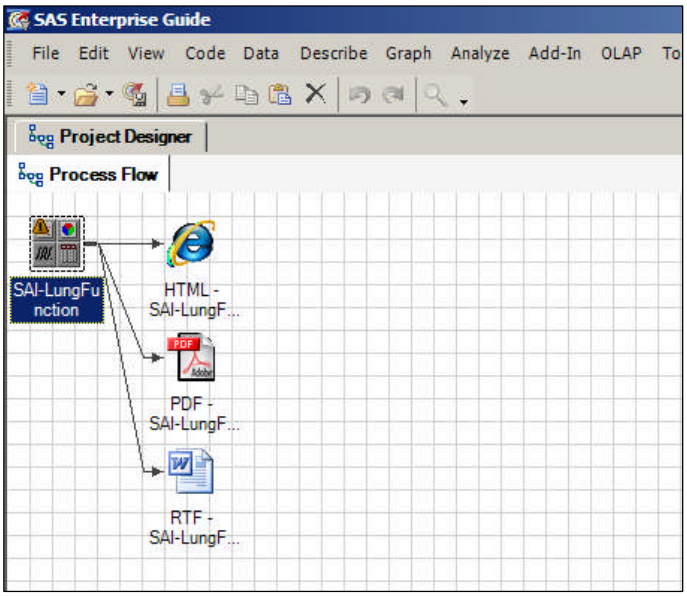

Figure 9

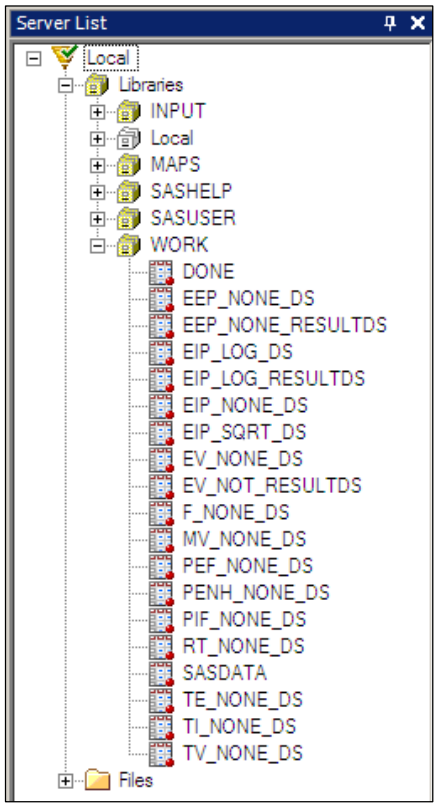

Figure 10

TECHNOLOGY USED

The Add-In application runs on a .Net framework (version 1.1) and consists of a user interface, which is developed in Visual Studio Professional 2003, and SAS (Version 9.1.3) programs which will analyse and process the data.
The user interface leads the user through the analyses he/she would like to perform and communicates with the SAS server on which the appropriate SAS programs will be run. The SAS server will return the results back to the user interface and/or Enterprise Guide.

The communication between the user interface and the SAS programs executed on the SAS server is mainly done with the help of %LET and %INCLUDE statements and SAS datasets. .NET will only be used to show results SAS has generated (datasets or parts of datasets) on the screen and to pass user requests and selections to SAS. All data processing, reporting and statistical analyses will be performed by SAS.
The list box containing the possible values of the 'Control group' on the user interface, for example, is filled with all items encountered in a specific column of a pre-specified SAS dataset. This pre-specified SAS dataset is created when the data is first imported and will contain all distinct values of the GROUP variable.

Whenever a user selects a button on a screen this will most of the time invoke one or more SAS programs. The selections the user makes in other screen elements that are appropriate for the SAS program will be passed to SAS before the program(s) is/are run. For example, when a user selects e.g. 'Normality and Homogeneity concluded with transformation:' the following will be passed to a SAS session:

```
%LET transval = [Value of the dropdown-list in the "Conclusion" section];
%LET parname  = [Name of current form; the parametername];
%LET control  = [Value of the dropdown-list in the "Control Group" section];
%LET positive = [Value of the dropdown-list in the "Positive Control Group"
                 section];
%LET statdose = [1 if 'Post-dose values versus Pre-dose values' is not checked.
                 Otherwise 0];
%INCLUDE global.sas; /*contains general settings*/
%INCLUDE normal_dunnett.sas; /*Performs ANOVA dunnett test*/
```

The global.sas and normal_dunnett.sas programs will use the previously defined macro variables in its code and after they have run, the results can be viewed with the appropriate preview button.

PROJECT METRICS

At the time this paper is written the Add-In application for Lung Function studies is installed at Notox and is in the OQ/PQ phase. Add-Ins for other type of studies are currently in the design phase and are expected to be implemented next year.

Next to the study director of Notox, three employees of OCS Consulting have been working on developing and programming the Add-In for Lung Function studies:

- A .NET programmer. Mainly responsible for the user interface and communication between the user interface and the SAS programs running on the SAS server;
- A SAS programmer/Project manager. Involved in the design of the application, programming of the SAS programs and managing the project;
- A statistician. Involved in the design of the application and advising the client. Also involved in programming of the statistical procedures within the SAS programs.

The design, development, documentation and validation of the Add-In for Lung Function application up to this point (without the OQ and PQ) has been completed in about 30 days.
These 30 days roughly consist of the following elements:

Design of the application:	5 days
Programming SAS code:	7 days
Creating the user interface:	13 days
Documentation and validation:	4 days
Project management:	1 day

Since the Lung Function application was the pilot project it is expected that the upcoming projects will take considerably less effort. The functionality of the elements of the user interface, for example selecting rows in a dataset shown within a data grid, does not need to be developed again and can be copied to other Add-Ins. Also, many of the available SAS programs can be used again with only minor adaptations.
Therefore it is expected that the development of an Add-In for another type of study will take approximately 6-8 days.

CONCLUSION

An Add-In for SAS Enterprise Guide provides users not only with a validated, structured environment in which he or she is guided through pre-defined decision steps, it also provides users the flexibility of performing exploratory analyses on validated imported study data, using all available SAS Enterprise Guide functionality.
All data handling, reporting and statistical analyses will be performed by SAS, which is broadly used and accepted within the pharmaceutical industry. The interaction between SAS and the user interface is not difficult to comprehend and therefore it is quite easy to make adaptations to the functionality of the Add-In.

The options selected at the various decision-points of the Add-In will be described in the generated reports and the SAS log-files will be available. Moreover the results of the ‘Add-In’ will be stored in the ‘Process Flow’ of SAS Enterprise Guide and can be saved within a project. Results and the way these results were obtained are therefore reproducible.

CONTACT INFORMATION

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

Author name	:	Bas van Bakel
Company	:	OCS Consulting
Address	:	PO BOX 490 5240 AL Rosmalen The Netherlands
Work phone	:	+31 (0)73 523 6000
Facsimile	:	+31 (0)73 523 6600
E-mail	:	sasquestions@ocs-consulting.com
Website	:	www.ocs-consulting.nl