

Using SAS® Enterprise Guide to Automate and Customize Statistical Analysis of Pharmaceutical Data

Alex Dmitrienko, Ph.D., Eli Lilly and Company, Indianapolis, IN

Cindy Lee, Eli Lilly and Company, Indianapolis, IN

ABSTRACT

With increasing demand to analyze massive clinical and non-clinical data, pharmaceutical companies are looking for tools to set up systems for automating and customizing statistical analysis and reporting. The systems reduce statisticians' time spent in programming and expedite the design, conduct and analysis of pharmaceutical studies. SAS Enterprise Guide provides a powerful graphical user interface that facilitates common tasks such as data access, analysis and reporting in a Windows environment. This presentation demonstrates the benefits of using SAS Enterprise Guide as a tool for creating analysis and reporting modules in pharmaceutical applications. As an illustration, the presentation describes custom Enterprise Guide tasks for designing clinical trials and summarizing clinical trial data.

INTRODUCTION

SAS Enterprise Guide is a relatively new SAS product that has already been through three major releases. Its latest version, Enterprise Guide 3.0, was rolled out early this summer. Enterprise Guide has attracted much attention in the SAS user community because it offers an intuitive point-and-click interface, provides the user with a large number of standard data manipulation/analysis tools and, last but not least, because it comes with a unique framework for developing custom tools.

The custom tools, known as custom Enterprise Guide tasks or add-ins, can be thought of as SAS macros with an easy-to-use graphical interface. Just like SAS macros, custom tasks input certain parameters, generate SAS code and create output or data sets. However, unlike regular SAS macros, custom tasks exhibit the following attractive features:

- Custom tasks are a lot more "intelligent" and user-friendly than SAS macros. The more complex the macro the longer it takes for the user to understand all of its options. Custom tasks can virtually eliminate the learning curve by walking the user through the process of performing a complex analysis or setting up a complex report. Additionally, custom tasks can run a variety of cross-checks along the way. As a result, custom Enterprise Guide tasks can greatly increase productivity and accelerate the design, conduct and analysis of pharmaceutical studies.
- Custom tasks are easy to distribute and upgrade (if desired, they can inform the user that an upgrade is available) and have a potential to replace quite a few expensive commercial software packages (the custom task for designing group sequential clinical trials introduced below serves as a good example).
- Lastly, since custom tasks are similar to SAS macros, they require the same level of validation; for example, they can be validated using validation procedures for multiple-use SAS macros. The advantage of using custom tasks is that they are tamper-proof (unlike SAS macros) and do not require any revalidation work once they have been distributed to users.

Custom tasks can be created using the Microsoft Component Object Model (COM) technology and can be written using Microsoft Visual Basic, C# or any other COM-compatible language. The custom tasks described in this presentation were written in Visual Basic 6.0 using sample code published on the SAS web site. The code was compiled as an ActiveX dynamic-link library (DLL) and the resulting custom tasks are compatible with both Enterprise Guide 2.0 and 3.0.

This presentation focuses on the use of custom Enterprise Guide tasks in the design of clinical trials and analysis of clinical trial data. The same approach can be applied to put together systems for automating and customizing statistical analysis/reporting in other areas of pharmaceutical industry as well as other industries.

CUSTOM TASK 1: DESIGN OF GROUP SEQUENTIAL CLINICAL TRIALS

Interim monitoring of safety and efficacy profiles of investigational drugs has become an integral part of clinical trials. Interim assessments are motivated by ethical requirements (it is imperative to ensure that patients are not exposed to harmful therapies) or financial considerations (to make optimal use of research and development dollars). Interim monitoring is performed in a group sequential manner, that is, interim data looks are taken after groups of patients have completed the study. A clinical trial is stopped as soon as it becomes evident that the investigational drug is superior or inferior to the control.

There are several commercially available software packages for designing group sequential clinical trials. They are rather expensive and, as all canned packages, cannot be customized. In this section we will introduce a custom Enterprise Guide task for designing group sequential trials. The task is based on the SAS macros published in Dmitrienko et al (2004, Chapter 4). See this book for more information about group sequential clinical trials and their implementation in SAS.

CLINICAL TRIAL EXAMPLE

Consider a clinical trial testing a single dose of a new drug for treating clinical depression versus placebo. The efficacy of the experimental drug will be evaluated using the mean change in the total score of the 17-item Hamilton Depression Rating Scale (HAMD17). The trial will have 80% power to detect a statistically significant treatment difference in the mean HAMD17 change at a two-sided 0.05 level assuming that the true mean treatment difference is 3 with a standard deviation of 8.

The clinical trial will employ two interim analyses and a final analysis. The interim looks will be taken after 50 and 75 percent of the patients have completed the study. The trial will be stopped early if the experimental drug is superior to placebo at either analysis.

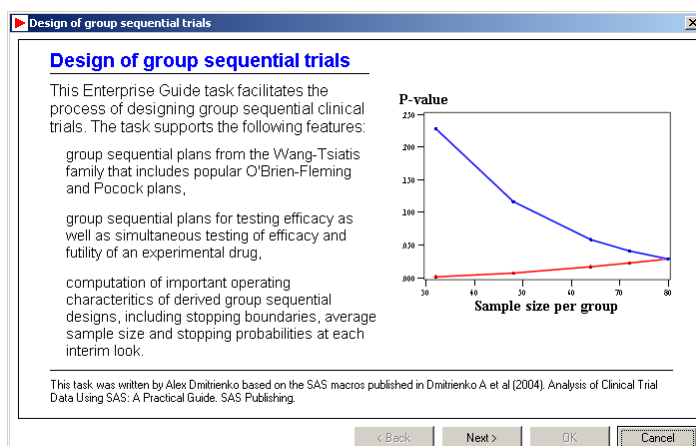
GROUP SEQUENTIAL DESIGN CUSTOM TASK

The group sequential design task will be utilized in order to compute important elements of the group sequential design, such as the maximum and average number of patients per treatment group. This task was built as a wizard that interviews the user and incorporates his/her responses into the SAS code it generates. Since the task is intended to be used at the design stage, it does not require an input data set.

Welcome Screen

The welcome screen of the custom task briefly summarizes the features the task supports and includes an example of a plot generated by the task (stopping boundaries of a group sequential design).

The user needs to click the Next button to proceed to Step 1.



Step 1

The user specifies the parameters of the group sequential design:

- One-sided Type I error rate (0.025),
- Power (0.8),
- Number of analyses (3),
- The timing of the analyses relative to the total sample size (0.5, 0.75 and 1),
- Expected treatment difference (3),
- Standard deviation (8).

The user needs to click the Next button to proceed to Step 2.

Design of group sequential trials

Step 1 Parameters of group sequential design

Overall Type I error rate and power

One-sided Type I error rate: 0.025

Note: To specify the usual two-sided Type I error rate of 0.05, enter a one-sided Type I error rate of 0.025

Power: 0.8

Number of analyses

Number of analyses (including the final analysis): 3

If the interim analyses are not equally spaced, specify the fraction of the total sample size for each analysis: 0.5, 0.75, 1.0

☐ Interim analyses will be equally spaced

Primary trial endpoint

☒ Continuous endpoint

Expected treatment difference: 3

Common standard deviation: 8

☐ Binary endpoint

< Back Next > OK Cancel

Step 2

The user specifies the objective of the clinical trial (stop due to overwhelming efficacy of the experimental drug) as well as the stopping boundary. The task provides information to help the user select the appropriate stopping boundary.

The user needs to click the OK button to run the task.

Design of group sequential trials

Step 2 Group sequential plan

Trial objective

☒ Stop early only in the presence of overwhelming efficacy

☐ Stop early due to overwhelming efficacy or futility

Shape parameter for efficacy stopping boundary: 0

Note: The shape parameter is a value between 0 and 0.5 that determines the shape of a stopping boundary. The smaller the value the less likely the study to stop early due to overwhelming efficacy.

Set the shape parameter = 0 to request the O'Brien-Fleming stopping boundary and 0.5 to request the Pocock stopping boundary. The O'Brien-Fleming boundary is commonly used for efficacy testing whereas the Pocock boundary is often used for futility testing.

< Back Next > OK Cancel

OUTPUT

Based on the responses provided by the user, the group sequential design task generates SAS code for computing important operating characteristics of the specified design. The output displayed in the Workspace window on the Enterprise Guide desktop includes a summary of the design's parameters (Table 1), stopping boundaries on test statistic and p-value scales (Table 2) and stopping boundaries under the null and alternative hypotheses, H0 and H1 (Table 3).

Table 1. Parameters of the group sequential design

Summary	Value
One-sided Type I error probability	0.025
Power	0.8
True effect size	0.38
Number of analyses	3
Fractions of total sample size	0.5, 0.75, 1.0
Shape parameter of efficacy boundary	0
Maximum sample size per group	112
Average sample size per group under H0	111
Average sample size per group under H1	109
Fixed sample size per group	109

Table 2. Stopping boundaries

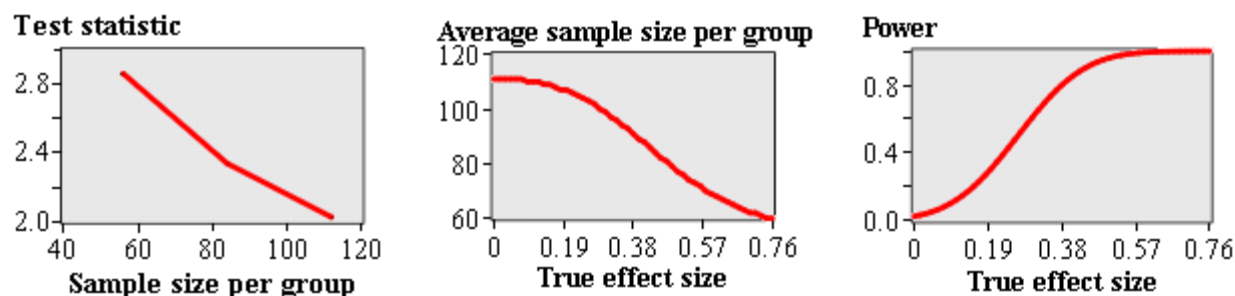
Analysis	Sample size per group	Stopping boundary (test statistic scale)	Stopping boundary (p-value scale)
1	56	2.8626	0.0021
2	84	2.3373	0.0097
3	112	2.0242	0.0215

Table 3. Stopping probabilities

Analysis	Sample size per group	Stopping probability under H0	Stopping probability under H1
1	56	0.0021	0.1954
2	84	0.0084	0.3567
3	112	0.0145	0.2479

The task also creates a series of plots to provide graphical summaries of computed stopping boundaries, power curve and average sample size as a function of the true effect size (Figure 1).

Figure 1. Stopping boundaries, power and average sample size of the specified design



CUSTOM TASK 2: SUMMARY OF PATIENT CHARACTERISTICS

While the custom Enterprise Guide task described in the previous section dealt with the design stage of clinical trials, this section introduces a custom task facilitating the analysis of clinical trial data and generation of publication-quality output.

CLINICAL TRIAL EXAMPLE

To demonstrate how to use this custom task, consider a SAS data set containing patient demographic data collected in a small clinical trial. There are six variables in the data set (Table 4). We are interested in producing a tabular summary of the demographic variables (gender, origin, age and age class) by dose and carrying out standard statistical tests to compare the distribution of the variables across the treatment groups.

Table 4. Patient demographics data set

Column heading	Description
Patient ID	Patient's identifier
Gender	Patient's gender ('M', 'F')
Origin	Patient's racial origin
Age	Patient's age in years
Age class	Age class (<65 years, ≥65 years)
Dose	Dose administered to patients (0 mg/day, 50 mg/day, 250 mg/day)

SUMMARY OF PATIENT CHARACTERISTICS CUSTOM TASK

The summary of the patient characteristics will be produced using a wizard-type task that goes through a series of screens to define row and column variables, descriptive statistics for numeric variables and statistical tests for comparing the demographic variables across the columns. The task is designed to be used at the analysis/reporting stage and thus it requires an input data set. The user must specify an input data set before launching the task.

Welcome Screen

The welcome screen of the task provides a summary of its features as well as an example of a summary table that can be generated using this task.

The user needs to click the Next button to proceed to Step 1.

Summary of patient characteristics

Summary of patient characteristics

This Enterprise Guide task generates tabular summaries of patient characteristics across treatment groups. The task supports the following features:

- summarizes both numeric and character variables,
- provides a large number of descriptive statistics for numeric variables,
- supports statistical comparisons across treatment groups,
- supports full customization of variable labels, variable formats, column headings and report title.

The task was written by Alex Dmitrienko

Variable	Dose (mg/day)	Comparison
Variable name	Statistic/Value	Placebo N=63 50 mg/day N=62 P-value
Age (years)	n	60 60
	Mean	55.1 57.8
	SD	11.1 10.2
Gender	F	45 (71.4%) 47 (75.8%) 686
	M	18 (28.6%) 15 (24.2%)

< Back Next > OK Cancel

Step 1

The user selects the row variables:

- Gender, origin and age class will need to be added as character variables,
- Age will be added as a numeric variable.

The user can use the Move up and Move down button to rearrange the row variables. Further, the user can modify the variable labels and formats.

The user needs to click the Next button to proceed to Step 2.

Summary of patient characteristics

Step1 Select row variables for the summary table

Variable list [PATINFO data set]

Variable	Label	Format	Length
PatientID	Patient ID		12
Gender	Patient's ...		8
Origin	Patient's ...		8
Age	Patient's ...		8
AgeClass	Age class		11
Dose	Dose (mg...		8

Add as a numerical variable
Add as a character variable

List of row variables

Name	Label	Format
Gender	Patie...	
Origin	Patie...	
Age	Patie...	
AgeClass	Age c...	

Variable properties

To change a variable's label or format in the summary table, select the variable and click Show Properties.

Show properties

Label

Format

Modify

Move up Move down

< Back Next > OK Cancel

Step 2

The user chooses the descriptive statistics for numeric variables.

The default descriptive statistics are:

- Number of patients with non-missing values,
- Mean,
- Standard deviation.

The user can rearrange the descriptive statistics and modify their labels.

The user needs to click the Next button to proceed to Step 3.

Summary of patient characteristics

Step 2 Select descriptive statistics for numerical row variables

Number and percentage of patients in each category will be computed for character variables

Summary of patient

Variable	Dose
Variable name	Statistic/Value
Age (years)	n 60
	Mean 55.1
	SD 11.1
Gender	F 45 (71.4%)
	M 18 (28.6%)

List of available descriptive statistics

Summary statistic	Label
Number of no...	n
Mean	Mean
Median	Median
Standard devi...	SD
Variance	Variance
Minimum	Minimum
Maximum	Maximum
Lower 95% co...	Lower 95% CL
Upper 95% co...	Upper 95% CL

Add >>

Selected descriptive statistics

Summary statistic	Label
Number of no...	n
Mean	Mean
Standard devi...	SD

<< Delete

Move up Move down

Statistic label

To change the selected statistic's label, press Show Label.

Show label

Label

Modify

< Back Next > OK Cancel

Step 3

The user needs to choose the column variable. In this case, the column variable is Dose with three values: 0 (Placebo), 50 (50 mg/day) and 250 (250 mg/day).

The user needs to click the Next button to proceed to Step 4.

Summary of patient characteristics

Step 3 Select column variable

Variable list [PATINFO data set]

Name	Label	Format	Length
PatientID	Patient ID		12
Gender	Patient's ...		8
Origin	Patient's ...		8
Age	Patient's ...		8
AgeClass	Age class		11
Dose	Dose (mg...		8

Define >>

Column variable

Dose

Header

Dose (mg/day)

Specify the number of columns and define column headers.

Number of columns

3

Column variable's value

0	Placebo
50	50 mg/day
250	250 mg/day

Column header

Placebo

50 mg/day

250 mg/day

< Back Next > OK Cancel

Step 4

Lastly, the user selects the statistical tests for comparing numeric or character variables across the three columns and specifies the table layout parameters (table width, title, footnote, etc).

The user needs to click the OK button to run the task.

Summary of patient characteristics

Step 4 Specify summary table parameters

Table characteristics

(mg/day)	Comparison
50 mg/day N=62	P-value
60	.168
57.8	
10.2	
47 (75.8%)	.686
15 (24.2%)	

P-values

☒ Compute p-values for across-column comparisons

Select tests for numerical and character variables

Numerical variables: F-test

Character variables: Exact chi-square test

Table layout

Table width (in): 7

Font size (pt): 12

Title: Summary of patient characteristics

Footnote: P-values for numerical variables are produced by the F-

< Back Next > OK Cancel

The custom task will generate a SAS program and submit it to the SAS System. After the program is executed, the following summary table will be displayed in the Workspace window:

Table 5. Summary of patient demographics

		Dose (mg/day)			
		0	50	250	
N		10	12	11	
Patient's gender					
F	n	7	10	9	
	%	70.0	83.3	81.8	
M	n	3	2	2	
	%	30.0	16.7	18.2	
Patient's origin					
AF	n	2	.	1	
	%	20.0	.	9.1	
CA	n	7	11	8	
	%	70.0	91.7	72.7	
HP	n	1	1	2	
	%	10.0	8.3	18.2	
Patient's age		Mean	55.4	55.5	56.7
		Std			
		Dev	11.8	13.7	8.2
Age class					
<65 yrs	n	7	9	8	
	%	70.0	75.0	72.7	
>=65 yrs	n	3	3	3	
	%	30.0	25.0	27.3	

It is important to note that, unlike built-in tasks available in Enterprise Guide 3.0, for example, Summary Tables task, the described custom task easily handles any number of numeric or character variables and supports statistical comparisons of row variables across the columns.

CONCLUSION

This presentation focused on the applications of advanced Enterprise Guide tools (user-written Enterprise Guide tasks) in the pharmaceutical industry. The tasks feature a point-and-click graphical interface and enable the user to gain access to the analytical power of the SAS software. Additionally, the tasks are self-documenting and can greatly improve productivity by eliminating the steep learning curve users typically face when dealing with complex SAS macros.

The powerful features of custom Enterprise Guide tasks were illustrated in this presentation using two tasks for pharmaceutical applications (design of group sequential clinical trials and analysis of clinical trial data). The tasks are implemented as ActiveX DLL files and can be downloaded from the Enterprise Guide user group's web site (<http://www.segus.org/customtasks>).

For more information about Enterprise Guide, see the SAS web site (http://www.sas.com/technologies/bi/query_reporting/guide) and Enterprise Guide user group's web site (<http://www.segus.org>).

REFERENCES

Dmitrienko A, Molenberghs G, Chuang-Stein C, Offen W. (2004). *Analysis of Clinical Trials Using SAS: A Practical Guide*. SAS Publishing, Cary, NC.

CONTACT INFORMATION

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

Alex Dmitrienko
Eli Lilly and Company
Lilly Corporate Center
Indianapolis, IN 46285
Work Phone: 317-277-1979
Fax: 913-402-7239
Email: Dmitrienko_Alex@lilly.com

Cindy Lee
Eli Lilly and Company
Lilly Corporate Center
Indianapolis, IN 46285
Work Phone:
Fax:
Email:

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.