

SAS Enterprise Guide in Pharmaceutical Applications: Automated Analysis and Reporting

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

ABSTRACT

SAS® Enterprise Guide® is a member of the SAS Business Intelligence suite, a broad collection of software products featuring intuitive point-and-click interfaces. Enterprise Guide offers an impressive selection of standard data manipulation and analysis tasks and provides a framework for developing new custom tasks. Due to an increasing demand to streamline routine data management and output generation tasks in the pharmaceutical world, Enterprise Guide will likely become one of the main tools for a broad group of pharmaceutical professionals involved in the so-called AAAA processes (data acquisition, aggregation, access and analysis), i.e., information technology, data management, statistics, etc. This paper describes how pharmaceutical companies can utilize Enterprise Guide to build systems for automating statistical analysis and reporting.

ENTERPRISE GUIDE CUSTOM TASKS

Enterprise Guide provides a powerful framework for developing self-contained modules known as custom tasks. The modules can be thought of as SAS macros with an attractive graphical user interface. Custom tasks are similar to SAS macros in that they input certain parameters, generate SAS code and create output or data sets. However, unlike regular SAS macros, custom tasks exhibit a number of useful features summarized below.

ENTERPRISE GUIDE CUSTOM TASKS AND LEARNING CURVE

Custom tasks are more “intelligent” and user-friendly than SAS macros. It is well known that powerful SAS macros are complex and create a steep learning curve even for experienced SAS users. This happens because SAS macros force the user to deal with each one of their numerous options or arguments. In contrast, custom tasks offer a user-friendly wizard-type interface that updates itself as the user makes his or her choices and reduces the number of arguments the user needs to specify. This feature helps eliminate the learning curve and enables less experienced users to quickly understand the process of performing a complex analysis or setting up a complex report. An additional advantage of using custom tasks is their ability to run cross-checks as the user makes choices or enters data. As a result, custom Enterprise Guide tasks can greatly increase productivity and accelerate the design, conduct and analysis of pharmaceutical studies.

DISTRIBUTION OF ENTERPRISE GUIDE CUSTOM TASKS

Custom tasks are easy to distribute and upgrade because each task is stored in a single file. If custom tasks are saved in a central location, individual users can copy them to their hard drives and install them using the Add-in Manager in Enterprise Guide. Once installed, custom tasks can be set up to search for their own updates and, if an update is found, inform the user. If desired, a custom task can help the user install a newer version.

VALIDATION ISSUES OF ENTERPRISE GUIDE CUSTOM TASKS

Finally, 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.

OVERVIEW

This paper focuses on the use of custom Enterprise Guide tasks in clinical trial applications. It introduces two custom tasks written by the author:

1. Multiple-Column Tabular Summary task. This task produces general-purpose multiple-column summary reports that can be used to display various types of clinical trial data. One example is the summary of patient demographic data by treatment group. This summary is used in virtually all clinical trials to examine the key characteristics of patients enrolled in the study and compare patient populations across the treatment groups.
2. Data Review task. This task facilitates the data quality review process that plays an important role in clinical trials (Brunelle and Kleyle, 2002). The task generates frequency counts for selected character variables and extreme values for selected numeric, date and time variables. A review of the produced frequency counts and extreme values helps the user identify inconsistent data values and data entry errors.

The two custom tasks were chosen to demonstrate the attractive features of Enterprise Guide tasks summarized above and give examples of how various groups of SAS users working in the area of clinical development (e.g., data management coordinators, statistical analysts and project statisticians) can benefit from this SAS product. It is worth noting that a similar approach can be applied to develop custom tasks for automating and customizing statistical analysis/reporting in other areas of pharmaceutical industry.

CUSTOM TASK DEVELOPMENT

Custom Enterprise Guide tasks are created using the Microsoft Component Object Model (COM) technology and can be written in Microsoft Visual Basic, C# or any other COM-compatible language. The custom tasks described in this presentation were written in Visual Basic .NET using sample code published on the SAS web site. For more information, see the Enterprise Guide page on the SAS web site (http://www.sas.com/technologies/bi/query_reporting/guide).

CUSTOM TASK 1: MULTIPLE-COLUMN TABULAR SUMMARY

Consider a custom task for creating a general-purpose tabular summary that compares multiple variables (both numeric and character) across treatment groups or important subgroups in clinical trials. Tabular summaries of this type are typically used for summarizing patient characteristics at the beginning of the study.

The Multiple-Column Tabular Summary task relies on a wizard-type interface to “interview” the user. The wizard goes through a series of screens to define the following parameters of the summary report:

1. Row variables (analysis variables).
2. Descriptive statistics for numeric row variables (frequency count and percentage will be computed for all character variables).
3. Column variable and column parameters.
4. Statistical tests to compare the row variables across the columns and also specification of table layout parameters.

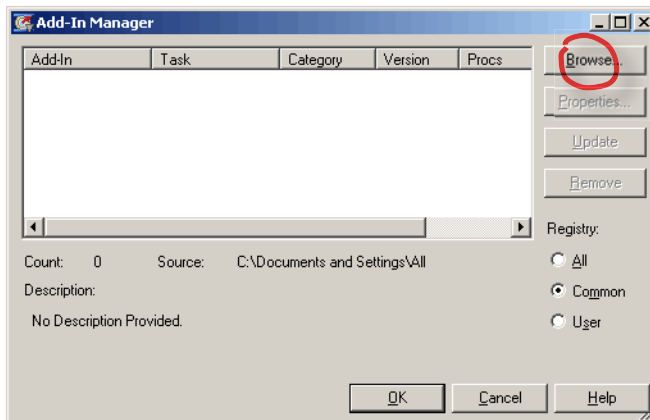
SUMMARY OF PATIENT DEMOGRAPHICS

As an illustration, consider a SAS data set containing patient demographic data from a clinical trial that was conducted to compare a single dose of an experimental drug to placebo. Suppose that you are interested in producing a summary of demographic variables (gender, origin, age and age class) by dose and carrying out statistical tests to compare the distribution of these four variables between the treatment groups. The six steps below illustrate the process of creating a summary of the four variables using the Multiple-Column Tabular Summary task. It will be assumed that you have Enterprise Guide 3.0 installed on your personal computer along with SAS Release 8.2 or later.

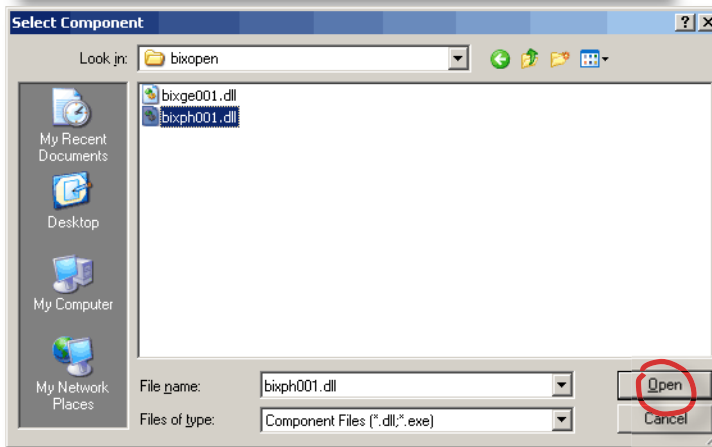
STEP 1: INSTALL THE MULTI-COLUMN TABULAR SUMMARY TASK

Before you can run the Multiple-Column Tabular Summary task, you need to install it on your computer. The task can be found in the Business Intelligence Exchange on the BISUG web site. To find the task, point your Internet browser to the Business Intelligence Exchange page (<http://www.bisug.org/bix/pharm.html>) and select the Multiple-Column Tabular Summary task. Save the task file (bixph001.dll) anywhere on your hard drive. Note that you must be logged in to your BISUG account to be able to download custom tasks from the Business Intelligence Exchange.

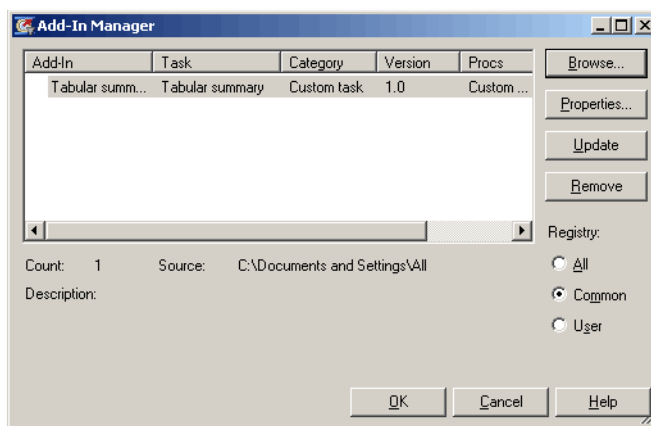
To make the downloaded file available to Enterprise Guide, choose the Add-In menu on the Enterprise Guide menu bar. This will bring up the Add-In Manager window displayed at right. Select Browse to tell Enterprise Guide where the task file is located.



Find bixph001.dll and click Open.



Enterprise Guide will install the Multiple-Column Tabular Summary task and add it to the list of available custom tasks.



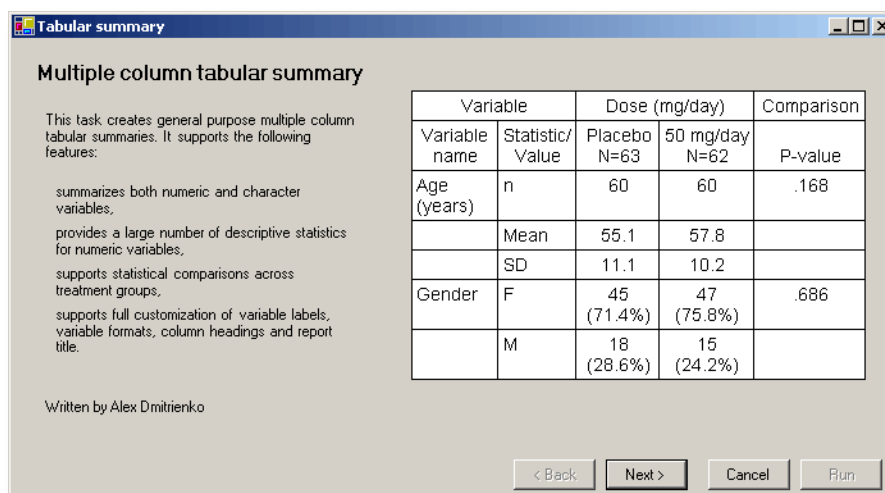
STEP 2: LAUNCH THE TASK

The Multiple-Column Tabular Summary task is designed to be used at the analysis/reporting stage and requires an input data set. To specify a data set, you can select a SAS data set on the Enterprise Guide desktop or a SAS data set residing on your hard drive or a network (you can download the sample data set used in this example from the BISUG website, <http://www.bisug.org/bix/pharm.html>). To launch the task, select the Add-In menu on the Enterprise Guide menu bar and choose the Multiple-Column Tabular Summary task.

As was noted above, the Multiple-Column Tabular Summary task features a wizard-type interface with a series of screens to interview the user.

The first screen provides a short description of the main features of the task and gives an example of a summary report it can generate.

Click Next to proceed to the next screen.

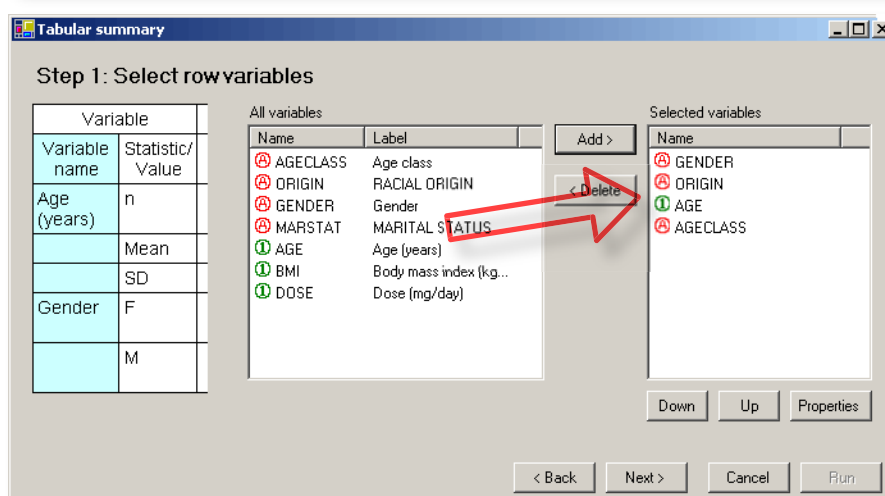


STEP 3: SPECIFY ROW VARIABLES

The next screen of the wizard enables you to choose the row variables that will be included in the summary report. To create a summary of the gender, origin, age and age class variables, select them in the left-hand pane and click Add. To specify multiple variables, press the Shift or Ctrl key and use the mouse pointer to select row variables.

Once the variables have been added to the right-hand pane, you can re-arrange them by using the Up and Down keys.

To modify their attributes (e.g., their labels), click Properties. Click Next to proceed to the next screen.



Note: The icons in the left- and right-hand panes help determine the variable type. The red icons identify character variables whereas the green icons identify numeric variables. No icon is displayed for other variable types (for example, date and time variables) because they cannot be included in the summary report.

STEP 4: SPECIFY DESCRIPTIVE STATISTICS FOR NUMERIC VARIABLES

This screen provides you with an option to select descriptive statistics for the numeric row variables. The following is a list of default descriptive statistics:

1. Number of non-missing values.
2. Mean.
3. Standard deviation.

To include other descriptive statistics, select them in the left-hand pane and click Add. There is no need to specify descriptive statistics for character row variables because they will be summarized using frequency counts and percentages. Click Next to proceed to the next screen.

Step 2: Select descriptive statistics for numeric row variables

Variable	Statistic/Value	Place	N=
riable	n		6
Mean	Mean		
Median	Median		
Standard deviation	SD		
Minimum	Minimum		
Maximum	Maximum		
Lower 95% confidence li...	Lower 95% CI		
Upper 95% confidence li...	Upper 95% CI		

All descriptive statistics

Statistic	Label
Number of non-missing ...	n
Mean	Mean
Median	Median
Standard deviation	SD
Minimum	Minimum
Maximum	Maximum
Lower 95% confidence li...	Lower 95% CI
Upper 95% confidence li...	Upper 95% CI

Selected descriptive statistics

Statistic
Number of non-missing values
Mean
Standard deviation

Buttons: Add >, < Delete, Down, Up, Properties, < Back, Next >, Cancel, Run

STEP 5: SELECT THE COLUMN VARIABLE

Here you need to specify the variable that defines columns in the summary report. In this case, Dose will serve as the column variable. Its label that will be included in the column heading is set to "Treatment group." Lastly, there will be two columns in the summary report. The first of them will include information on records with Dose=0 and will be labelled "Placebo." The other column will correspond to Dose=50 and will be labeled "50 mg/day." Click Next to proceed to the last screen.

Step 3: Select the column variable

Dose (mg/day)	
Placebo N=63	50 mg/day N=62
60	60
55.1	57.8
11.1	10.2
45 (71.4%)	47 (75.8%)
18 (28.6%)	15 (24.2%)

Column variable

Column variable: DOSE

Header: Treatment group

Columns

Number of columns: 2

Column variable's value	Column header
0	Placebo
50	50 mg/day

Buttons: < Back, Next >, Cancel, Run

STEP 6: DEFINE STATISTICAL TESTS AND TABLE LAYOUT PARAMETERS

The last screen of the wizard contain two panes. The left-hand one displays available statistical tests for numeric and character variables. The right-hand pane enables you to customize the table layout, for example, specify the font size and report's title and footnote.

To generate a summary of patient characteristics, click Run.

Step 4: Specify summary table parameters

P-values

☒ Compute p-values for across column comparisons

Test for numeric variables: F test

Test for character variables: Chi square test

Table layout

Table width (in): 7

Font size (pt): 12

Title: Summary of patient characteristics

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

Buttons: < Back, Next >, Cancel, Run

After you click the Run button, the Multiple-Column Tabular Summary task will generate a SAS program and submit it to the SAS System installed on your computer. The summary report produced by the program will be displayed in the Workspace window. The summary of patient characteristics displayed below shows that the selected demographic variables are comparable between the two treatment groups.

RTF - Summary of patient characteristics

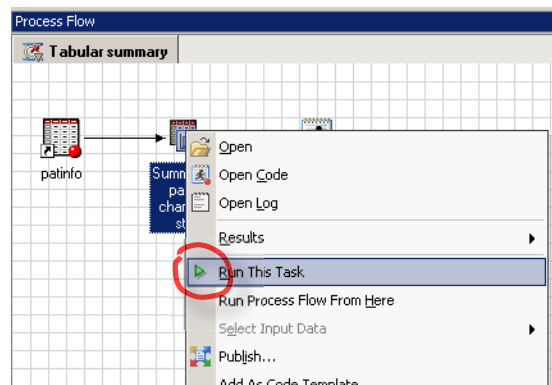
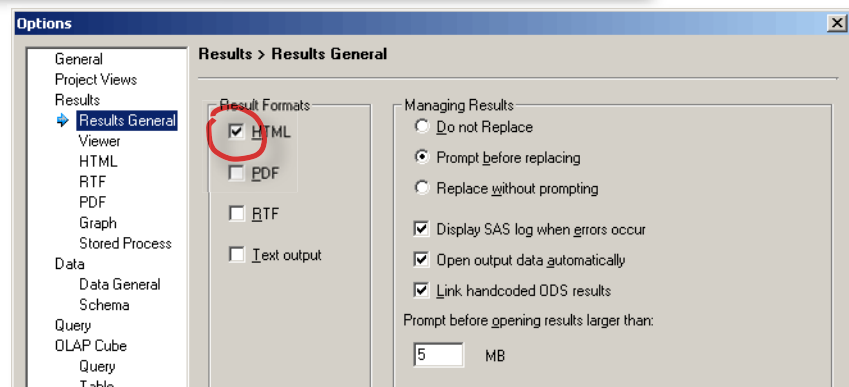
Summary of patient characteristics

Variable		Dose (mg/day)		Comparison
Variable name	Statistic/Value	Placebo N=63	50 mg/day N=62	P-value
Gender	F	45 (71.4%)	47 (75.8%)	.686
	M	18 (28.6%)	15 (24.2%)	
Racial origin	AF	3 (4.8%)	0 (0.0%)	.362
	CA	56 (88.9%)	59 (95.2%)	
	HP	3 (4.8%)	3 (4.8%)	
	O	1 (1.6%)	0 (0.0%)	
Age (years)	n	60	60	.168
	Mean	55.133	57.833	
	SD	11.118	10.185	
Age class	.	3 (4.8%)	2 (3.2%)	.536
	<65 yrs	46 (73.0%)	40 (64.5%)	
	>=65 yrs	14 (22.2%)	20 (32.3%)	

Note that in this case, the report was generated in a rich text format (RTF).

To change the output format, you will need to go through the following steps:

1. Select the Tools menu on the Enterprise Guide menu bar and then choose Options.
2. When the Options window appears on the screen, select the Results General category and check the appropriate box or boxes, for example, HTML.
3. Click OK to save the output options.
4. Re-run the Tabular Summary task by right-clicking its icon in the project window and selecting Run This task.



CUSTOM TASK 2: DATA QUALITY REVIEW

The second example involves a custom task that facilitates data quality reviews. It is well known that the clinical trial sponsor is responsible for ensuring the accuracy and completeness of the clinical trial database. Data reviews are routinely used in clinical trial to assess the data quality and identify spurious data values or data collection/data entry errors. These data reviews can be conducted in a systematic manner in the context of the so-called database quality reviews or in a more informal fashion by examining the distribution of important variables.

An efficient way of identifying problems in clinical trial data is based on a review of frequency counts for important character variables and extreme (smallest and largest) values for important numeric, date and time variables. By examining frequency counts, one can assess the prevalence of missing values or data entry errors (e.g., misspelled words). Likewise, a review of extreme values helps identify data points that are not consistent with the rest of the data (outliers) and most likely represent data collection or entry errors.

The Data Review task provides a useful tool that facilitates a review of the distribution of important variables. It is similar to the Multiple-Column Tabular Summary task in that it also relies on a wizard-type interface. To specify the parameters of the data quality review report, the user needs to go through the following three screens:

1. Specification of the ID variable that uniquely identifies each record in the data set.
2. Selection of review variables for which a frequency count or extreme values need to be generated.
3. Selection of secondary variables that help examine the distribution of the review variables.

REVIEW OF BASELINE PATIENT DEMOGRAPHICS

To demonstrate how the Data Review task can be used in a clinical trial setting, consider first a patient demographics data set. The data set contains one record per patient.

	PATIENT	AGECLASS	ORIGIN	GENDER	EDUCAT
1	101	<65 yrs	CA	M	12
2	101	<65 yrs	CA	M	12
3	101	<65 yrs	CA	M	12
4	101	<65 yrs	CA	M	12
5	102	>=65 yrs	CA	F	13
6	102	>=65 yrs	CA	F	13
7	102	>=65 yrs	CA	F	13
8	102	>=65 yrs	CA	F	13
9	103	<65 yrs	CA	F	99
10	103	<65 yrs	CA	F	99
11	103	<65 yrs	CA	F	99
12	103	<65 yrs	CA	F	99
13	104	<65 yrs	HP	F	99
14	104	<65 yrs	HP	F	99
15	104	<65 yrs	HP	F	99
16	104	<65 yrs	HP	F	99

STEP 1: INSTALL THE DATA REVIEW TASK

To install the Data Review task, download the task file (bixge001.dll) from the BISUG web site (<http://www.bisug.org/bix/general.html>) as described earlier.

STEP 2: LAUNCH THE DATA REVIEW TASK

The Data Review task requires an input data set and thus you need to select a data set in the Project window before launching the task or browse your hard drive to specify the input data set after launching the task. Also, as before, to run the task, select the Add-In menu on the Enterprise Guide menu bar and choose the Data Review task.

The first screen of the Data Review task describes the main features of this task and provides an example of a report it generates.

The screen also shows you what variables need to be specified to produce a data quality review report. The Subject variable is as a good example of an ID variable since it uniquely identifies each record in this data set. The Weight variable serves as a review variable and the reports lists its three smallest values. Lastly, the Height and Gender variables are (optional) secondary variables that will help you determine whether or not the three weights are feasible.

Click Next to proceed to the next screen.

Subject	Weight	Height	Gender
293	110	61	F
052	111	72	M
131	114	70	M

ID variable Review variable Secondary variables

< Back Next > Cancel Run

STEP 3: SELECT THE ID VARIABLE

This screen includes two panes. In the left-hand pane you need to select the ID variable from a dropdown list of all variables in the selected data set (in this case it is the Patient variable). The right-hand pane shows the number of extreme values that will be listed for each numeric, date and time review variable. The default value is 5 and it generally makes sense to select less than 10 extreme values. Click Next to proceed to the next screen.

The dialog box is titled "Data quality review". It has two main sections: "Select ID variable and data summary options".

On the left, there is a list of variables: "Subject", "293", "052", "131".

On the right, there is a section titled "Options". It contains a dropdown menu for "Select ID variable" with "PATIENT" selected. Below it, a text box says "ID variable uniquely identifies each record in the data set". To the right of this, there is a section titled "Options" with a dropdown menu for "Select the number of extreme values" with "3" selected. Below it, a text box says "3 smallest and largest values will be included in the data summary for each selected numeric, date or time variable".

At the bottom, there are four buttons: "< Back", "Next >", "Cancel", and "Run".

STEP 4: SELECT REVIEW VARIABLES

This screen enables you to specify review variables that will be included in the data quality review report. To choose the Weight variable, select it in the left-hand pane and click Add.

If you would like to select multiple variables, press the Shift or Ctrl key and use the mouse pointer to select variables in the left-hand pane.

To generate data quality review report containing all of the variables, click the Add all button.

Click Next to proceed to the next screen.

The dialog box is titled "Data quality review". It has two main sections: "Select review variables".

On the left, there is a list of variables: "Weight", "110", "111", "114".

On the right, there is a section titled "Review variables". It contains a table with two columns: "Name" and "Label". The table has the following rows: "PATIENT", "AGECLASS", "ORIGIN", "GENDER", "EDUCAT", "MARSTAT", "DURDISCL", "WEIGHT", "HEIGHT". The "WEIGHT" row is selected. To the right of the table, there are buttons: "Add", "Delete", and "Add all >>".

At the bottom, there are four buttons: "< Back", "Next >", "Cancel", and "Run".

STEP 5: SELECT SECONDARY VARIABLES

This is an optional step that deals with the so-called secondary variables. As was explained earlier, secondary variables help identify feasible values of review variables. For example, as shown on the first screen, it is prudent to display height and gender of patients with a very low weight.

To add the Height and Gender variables as secondary variables, select the Weight variable from the dropdown list and then choose Height and Gender from the list of secondary variables (you will need to press the Ctrl key to select the two variables simultaneously). After that click the Add button and the Data Review task will add two rows to the right-hand pane to indicate that the data quality review report for the Weight variable will include patients' height and gender.

The dialog box is titled "Data quality review". It has two main sections: "Select secondary variables".

On the left, there is a list of variables: "Height", "Gender", "61", "72", "70".

On the right, there is a section titled "Review and associated secondary variables". It contains a table with two columns: "Review" and "Secondary". The table has the following rows: "WEIGHT", "GENDER", "WEIGHT", "HEIGHT".

At the bottom, there are four buttons: "< Back", "Next >", "Cancel", and "Run".

To generate the data quality review report for the Weight variable, click Run. The SAS code generated by the Data Review task will be submitted to the SAS System and the output will be displayed in the Workspace window.

The resulting data quality review report lists the three smallest and three largest values of the Weight variable. The report also displays Patient IDs to make it easier to generate queries if the listed patient weights appear problematic.

A quick review of the report reveals that the Weight values are generally consistent with the height values except for Patient 1703 who is 63 in tall yet weighs a little over a hundred pounds. The patient's weight or height could have been entered incorrectly and it is prudent to check the database records for this patient against her case report form.

patient	weight	height	gender
1703	104	63	Female
1004	109	51	Female
104	111	54	Female
2804	280	64	Female
2907	287	68	Male
1202	290	67	Male

REVIEW OF EFFICACY DATA

Now consider a data set that contains efficacy data from a rheumatoid arthritis trial. The efficacy profile of rheumatoid arthritis therapies is often evaluated by studying their effect on patient's pain assessment as well as patient's global assessment of disease activity. These assessments are performed using a 100-mm scale. A score of 0 on the pain scale indicates no pain whereas a score of 100 indicates the patient is in a severe pain. A patient's global assessment of disease activity is performed in a similar manner.

Suppose that the patient pain assessment variable (PtPain) is your main focus and therefore it will serve as the review variable.

The patient pain and global assessment are known to be correlated and it is reasonable to add the patient global assessment variable (PtGlobal) as a secondary variable. Secondly, since the experimental therapy is likely to relieve the patients' pain and thus the pain and global assessment are likely to change over time, the visit variable (Visit) can provide useful information as a secondary variable.

Data quality review

Select secondary variables

For each review variable, you can choose one or more secondary variables that will be included in the data summary

Review variable: PTPAIN

Secondary variables:

Name
PATIENT
VISIT
PTPAIN
PTGLOBAL

Review and associated secondary variables

Review	Secondary
PTPAIN	VISIT
	PTGLOBAL

< Back Next > Cancel Run

The report generated by the Data Review task displays the three smallest and three largest values of the PtPain variable accompanied by patient's ID, patient global assessment and visit number. The patient pain and patient global assessments appear consistent for Patients 312, 404, 803 and 602. In the other two cases (Patients 2601 and 3105), the global score is considerably larger than the pain score and it will be advisable to verify these scores by comparing the database records to the case report forms.

patient	ptpain	ptglobal	visit
312	5	10	3
2601	7	34	2
3105	7	49	1
404	97	66	2
803	97	72	1
602	98	78	2

LEARN MORE ABOUT ENTERPRISE GUIDE

If you are interested in learning more about Enterprise Guide, please see A Visual Guide to Enterprise Guide. This electronic book provides a concise review of main features of Enterprise Guide and is an ideal way to get up to speed with this powerful software package. For more information about the book, please visit the Business Intelligence SAS Users Group (BISUG)'s web site:

<http://www.bisug.org/visualguide>

Sign up for the BISUG membership to receive access to information on SAS Business Intelligence products including free Enterprise Guide tutorials, papers presented at recent SAS conferences, review of Enterprise Guide training courses and custom tasks submitted by Enterprise Guide users.

CONCLUDING REMARKS

The paper introduced custom Enterprise Guide tasks and demonstrated how they can be utilized to facilitate the review and analysis of clinical trial data. The custom tasks feature a point-and-click graphical interface and enable the user to gain access to the analytical power of the SAS software. The tasks are also self-documenting and can greatly improve productivity of Enterprise Guide users.

For more information about Enterprise Guide and custom Enterprise Guide tasks, see

1. SAS web site (http://www.sas.com/technologies/bi/query_reporting/guide),
2. Business Intelligence SAS Users Group's web site (<http://www.bisug.org>).

Sign up for the Business Intelligence SAS Users Group (BISUG) membership to receive access to information on SAS Business Intelligence products including free Enterprise Guide tutorials, papers presented at recent SAS conferences and custom tasks submitted by Enterprise Guide users.

REFERENCES

Brunelle R, Kleye R. A database quality review process with interim checks. Drug Information Journal. 2002; 36: 357-67.

CONTACT INFORMATION

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

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.