

Delivering Statistics - A Workflow Supported by SAS Stored Processes

Volker Harm, Bayer Schering Pharma AG, Berlin, Germany

ABSTRACT

One way of delivering statistics to our scientist is to provide applications that enable the evaluation of standard tasks. The in vivo micronucleus test (MNT) is a test used in toxicological screening for potential genotoxic compounds. This test is routinely done in different locations. It needs statistical evaluation, which is too complex to be done by the scientist.

We have implemented a new workflow supported by a SAS® Stored Process, where the scientist starts the evaluation by providing the data and starting the evaluation from the intranet. The stored process performs the statistical analysis and creates different result objects in RTF to be included in reports as well as a listing containing the statistical results and delivers them to the statistician for approval. Thus we can substantially reduce the effort to give statistical support.

INTRODUCTION

THE IN VIVO MICRONUCLEUS TEST (MNT)

The OECD Guideline gives the following description of the in vivo micronucleus test:

"The micronucleus assay is an in vivo cytogenetic test which uses erythrocytes in the bone marrow of rodents to detect chemical damage to the chromosomes or mitotic apparatus of mammalian cells. As the erythroblast develops into an erythrocyte (red blood cell), its main nucleus is extruded and may leave a micronucleus in the cell body; a few micronuclei form under normal conditions in blood elements. This assay is based on an increase in the frequency of micronucleated erythrocytes found in bone marrow from treated animals compared to that of control animals. The visualization of micronuclei is facilitated in these cells because they lack a main nucleus...

Micronuclei mean small particles consisting of acentric fragments of chromosomes or entire chromosomes, which lag behind at anaphase of cell division. After telophase, these fragments may not be included in the nuclei of daughter cells and form single or multiple micronuclei in the cytoplasm.

Polychromatic erythrocyte (PCE) means an immature red blood cell that, because it contains RNA, can be differentiated by appropriate staining techniques from a normochromatic erythrocyte (NCE), which lacks RNA. In one to two days, a PCE matures into a NCE."

THE STATISTICAL TASK

The strategy for statistical analysis of the in vivo micronucleus test is based on significance tests for differences between mean numbers of micronucleated polychromatic erythrocytes (PCEM), micronucleated normochromatic erythrocytes (NCEM), and the ratio of polychromatic to normochromatic erythrocytes (PN) in control and treated samples.

THE WORKFLOW

The MNT is conducted by our Genetic Toxicology group. The data resulting from that experiment is sent to our group consisting of an Excel data sheet containing general information about the experiment and table of results in ASCII format. In our group a macro is executed for analysis of the data. The output is then interpreted by a statistician. The result is a set of means that show statistically significant differences. The descriptive statistics of the program run are read into a Microsoft Word document, and the significances are inserted manually as stars. This is cumbersome and error prone. The resulting report is then sent to the scientists via email. There all the result documents are compiled into a research report.

This workflow is replaced by a stored process application, which is executed by the scientists. Data describing the experiment are entered interactively from the web browser, measurements are read from specified location on a server, and the stars are automatically created according to rules given by our statisticians. The detailed statistical analysis output is produced in addition for checking and approval by the statistician.

DATA TO BE ANALYZED

EXPERIMENTAL DESIGN

The MNT is carried out with the following experimental design:

Group	Treatment	Dose level	24h	48h
1	Vehicle control		+	+
2	Test substance	Low dose	+	-
3	Test substance	Mid dose	+	-
4	Test substance	High dose	+	+
5	Positive control		+	-

The study method has a repeated measures parallel design, with two measurement time points (24h and 48h after administration). At the second time point (48h), only the vehicle control and the high dose group are measured. Male or female or both sexes are used in a study. Usually five animals per sex are included into a study.

MEASUREMENTS

The following study parameters are measured during the study:

Parameter	Description
PCE	Number of polychromatic erythrocytes per 1000 NCEs
NCE	Number of normochromatic erythrocytes
PCEM	Promille of micronucleated PCEs calculated from 2000 PCEs
NCEM	Promille of micronucleated NCEs calculated from 2000 NCEs

STATISTICAL ANALYSIS

ENDPOINTS

The following questions are investigated:

- Is there an increase of PCEM in the test dose groups and in the positive group compared to the vehicle control group at any of the time points.
- Is there an increase of NCEM in the test dose groups and in the positive group compared to the vehicle control group at any of the time points.
- Is there a change of the PN ratio in the test dose groups and in the positive group compared to the vehicle control group at any of the time points.

For the statistical analysis of these questions the ratio $PN = PCE/NCE$ of polychromatic and normochromatic erythrocytes is calculated, and the PCEM and NCEM values are transformed by calculating the arcsine of the root of these values.

Thus we get the following three endpoints for the statistical analysis:

- transformed PCEM
- transformed NCEM
- PN:

PROCEDURE

The following procedure is used for the evaluation of data:

For each study endpoint the positive control group is compared to the vehicle control group using means tests of general linear models (t-tests).

After that, when both sexes are used in the study, a general model with the treatment, (without the positive control) the gender and their interaction is fitted separately for the time points of 24h and 48h, and in both cases means tests for treatment and gender are calculated.

Also, the following contrasts are analyzed:

- Vehicle control vs. low dose, male animals
- Vehicle control vs. mid dose, male animals
- Vehicle control vs. high dose, male animals
- Vehicle control vs. low dose, female animals
- Vehicle control vs. mid dose, female animals
- Vehicle control vs. high dose, female animals

When only one sex is used in the study, then a general model with only the treatment (without the positive control) as independent variable is applied, and means tests for treatment are calculated.

Also, the following contrasts are analyzed:

- Vehicle control vs. low dose
- Vehicle control vs. mid dose
- Vehicle control vs. high dose

The tests are performed as two sided. The significance level is set to 5%.

Mean values and standard deviations are calculated for each treatment group and each time point separately for each sex and for the pooled data across genders. When sex dependent differences are found, the separate descriptive statistics are reported. When no sex dependent differences are found, then only the descriptive statistics of pooled data are used in the reports.

THE SAS PROGRAM

THE EXISTING MACRO

The procedure described above is implemented in a macro which was used in our department to evaluate the MNT for over a decade. It reads in the measurements, calculates the descriptive statistics with proc univariate and performs the tests with proc glm. Furthermore it produces a customized report that is really well established within the scientists' departments. What causes nuisance each time the macro is used in our department, is the manual evaluation of the proc glm listings, because proc glm produces a vast amount of output that you have to search for the required p-values and because manually marking of single values in a report is extremely error prone.

EXTRACTING P-VALUES WITH ODS

For the new macro we use the possibility with ODS to deliver output to data sets (see Olinger, C.R., Tobias, R. D.).

The idea is to create a data set with "stars" that can be merged to the descriptive statistics.

Using the ods output statement it is possible to select the tables that contain the p-values in proc glm. The following code shows a typical case:

```
proc glm data = Endpoints48h order = internal;
  class treat sex;
  model &Endpoint. = treat sex treat*sex;
  means treat sex/lsd;
  contrast "Plac. vs. Dosis3, M" treat -1 1 treat*sex -1 0 1;
  contrast "Plac. vs. Dosis3, F" treat -1 1 treat*sex 0 -1 0 1;
  ods output OverallANOVA = OverallANOVA;
  ods output ModelANOVA = ModelANOVA;
  ods output Contrasts = Contrasts;
run;
```

With these three ods output statement we create three compact data sets that contain the p-values with following typical structure:

Obs	Dependent	Source	DF	SS	MS	FValue	ProbF
1	pn	Plac. vs. Dosis3, M	1	0.07992360	0.07992360	6.39	0.0224
2	pn	Plac. vs. Dosis3, F	1	0.03516490	0.03516490	2.81	0.1130

Stepping through statistical analysis procedure as described above information is added to these table that is needed for merging it later to the descriptive statistics as well as the possible star. This is done with following statements:

```
data Contrasts;
  set Contrasts;
  if ProbF le 0.05 then star = '*';
  time = 2;
  treat = 4;
  sex = 3;
  _LABEL_ = "PCE/NCE";
run;
```

This adds the following columns to the above table:

Obs	star	time	treat	sex	_LABEL_
1	*	2	4	3	PCE/NCE
2		2	4	3	PCE/NCE

Appending these datasets along the statistical procedure we get a table stars that can be merged to the descriptive statistics. The existing code for the production of the creation of output data sets could be easily adapted.

PhUSE 2008

DIRECTING THE OUTPUT

With this toolset it was a pleasure to prepare the new workflow by directing the output to the various destination, as demonstrated by the following macro, which is a wrapper for the stored process:

```
%macro EvaluateMNT;  
    %Close_webout;  
    %SetEnvironment;  
    %OpenAnalysisFile;  
    %ImportData;  
    %PrepareDataForAnalysis;  
    %DoStats;  
    %CloseAnalysisFile;  
    %CreateOutputDSS;  
    %OpenReportFile;  
    %ReportResults;  
    %CloseReportFile;  
    %Open_webout;  
    %SayFarewell;  
%mend;
```

As a default, the stored process writes to the destination _webout. As the program produces a lot of output, the first step is to close this interactive destination. After setting the environment, such as libraries used, input file names etc. the traditional listing output file is opened. The next steps contain the statistical analysis. All output generated during these steps is written to this "analysis file".

For the output of the report file an ods rtf destination is opened. In this way the results produced by proc report can directly be imported into the scientists' research reports.

At the end _webout is opened again and links to the result files are provided.

THE NEW WORKFLOW

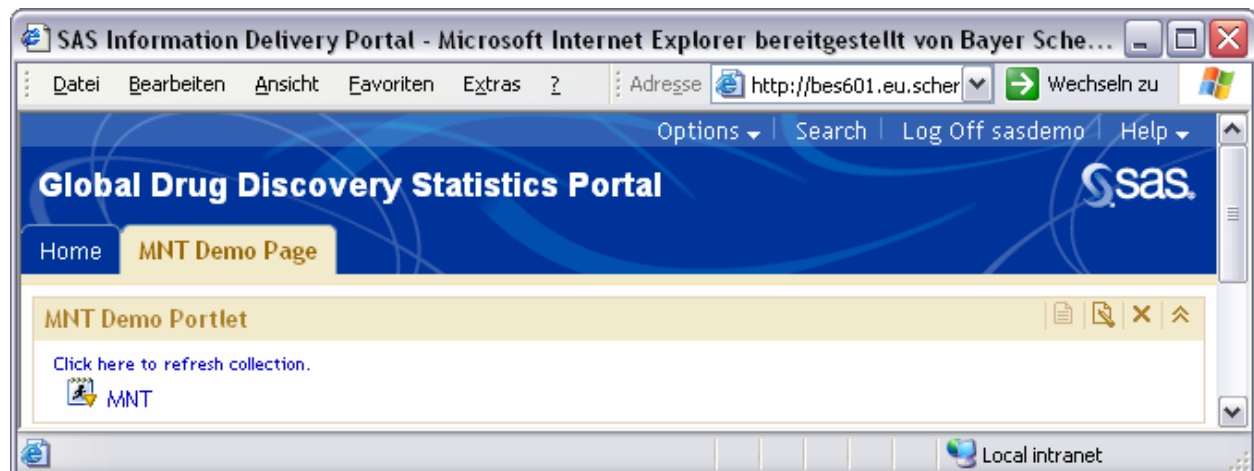
THE SAS STORED PROCESS

The new SAS program is converted to a SAS stored process and registered in the SAS Metadata via SAS Management Console. Thus it can be accessed from any web browser using the SAS® Stored Process Application or it can be integrated into the SAS® Information Delivery Portal.

In our department we build up a Global Drug Discovery Portal that provides statistical applications like this for the scientists in drug discovery.

To evaluate an MNT the measurements recorded in ASCII file have to be placed in a prespecified location, which can be identified by a project number and a study number.

When this is done the stored process can be started from a portlet within the Global Drug Discovery Statistics Portal:



PhUSE 2008

PARAMETERS

When the stored process is started, a window pops up that show the entry mask for the parameters needed to start the evaluation of the experiment.

The screenshot shows a web browser window with the address bar displaying `http://bes601.eu.schering.net:8080/Portal/viewItem.do?com.sas.portal.ItemId=Stored...`. The browser's menu bar includes 'Datei', 'Bearbeiten', 'Ansicht', 'Favoriten', 'Extras', and '?'. The address bar shows 'Adresse' and the URL `http://bes601.eu.sc`, with a 'Wechseln zu' button. The page header features 'Portal Page', 'Search', 'Help', and the 'sas' logo. Below the header is a navigation bar with 'Bookmark', 'Publish', and 'E-mail' links. The main content area is titled 'Parameters' and contains a table with the following entries:

Parameters	
* Project number:	EFGH
* Study number:	TXSTxxxxxxxx
* Test substance:	ZK ABC
* Vehicle control dose:	20ml/kg
* Test substance - low dose:	300mg/l
* Test substance - mid dose:	600mg/kg
* Test substance - high Dose:	1200ml/kg
* Positive control dose:	30mg/kg
Application route:	Group 1 - 4 i.v., Group 5 i.

Below the table, a note states: '* Indicates a required field.'

On the left side of the 'Parameters' section, there is a blue sidebar with the text 'Execution Options' and an 'Execute' button. The bottom of the browser window shows a taskbar with 'Parameters' and 'Local intranet' icons.

Project number and study number uniquely identify the experiment. From this unique identification the location for the measurements file can be derived. Test substance gives the name of the substance to be analyzed. Furthermore dose values for the different groups have to be entered as well as a description of the application route. The dose values are used to define a format for the dose groups.

PhUSE 2008

EXECUTION OPTIONS

There is a second section of parameters called "Execution Options" provided by SAS. This options provided there can be limited to the following three:

The screenshot shows a web browser window with the URL `http://bes601.eu.schering.net:8080/Portal/viewItem.do?com.sas.portal.ItemId=Stored...`. The page has a blue header with the MNT logo and a SAS logo. Below the header is a navigation bar with links for Portal Page, Search, Help, Bookmark, Publish, and E-mail. The main content area is divided into two sections: Parameters and Execution Options. The Execution Options section contains three rows of options:

Execution Options	
Display results in a new window?	☒ Yes ☐ No
HTTP method:	☐ Get ☒ Post
Process blank parameter values?	☐ Yes ☒ No

Below the Execution Options section is a large blue button labeled "Execute".

THE RESULTS WINDOW

The statistical analysis is started by clicking on "Execute". The stored process does the evaluations, creates the report for the scientist in RTF format and the traditional listing output for the statistician. A new output window will open showing a note that a run was executed and links to the result files.

The screenshot shows a web browser window titled "SAS Output - Microsoft Internet Explorer bereitgestellt von Bayer Schering Pharma BLN". The address bar shows `http://bes601.eu.scher...`. The main content area displays the following text:

MNT - Evaluation of the In Vivo Micronucleus Test
Run from 28AUG08 - 16:35
The following result file were created:

Below the text is a box containing a "Link" button and two links: "Analysisfile" and "Reportfile".

PhUSE 2008

THE ANALYSIS FILE

The analysis file contains the traditional SAS listings output. This output can be used to trace the analysis and besides the statistical evaluation it contains listings with the measurements as imported from the ASCII file, the endpoint values derived from the measurement as well as tables of descriptive statistics.

THE REPORT FILE

The report file is an RTF file containing a results table for each dose and a summary table for the whole experiment. These tables can be directly imported into the research report which are create in Microsoft word. As an example the summary table is shown below:

Table 6: Summary of the results:

Frequency of micronucleated polychromatic erythrocytes (in per mill) based on 2000 cells scored per animal and ratio of PCE/NCE in mouse bone marrow after Group 1-4: i.v. Group 5: i.g. of XX 12-707

Sampling times: 24 hours and 48 hours after treatment

Summary of results from table 1-5

Time of measurement	Treatment	Dosage	Number of animals	Male	Female	Male and Female	
				PCE(M) (MV +/- SD)	PCE(M) (MV +/- SD)	PCE(M) (MV +/- SD)	Ratio ¹ PCE/NCE
24 h after treatment	Vehicle control	10 ml/kg	5M/5F	1.00 +/- 0.50	0.90 +/- 0.42	0.95 +/- 0.44	0.92 +/- 0.09
	XX 12-707	2.5 mg/kg	5M/5F	1.20 +/- 1.30	0.70 +/- 0.27	0.95 +/- 0.93	0.94 +/- 0.11
	XX 12-707	5.0 mg/kg	5M/5F	0.70 +/- 0.27	0.60 +/- 0.22	0.65 +/- 0.24	0.90 +/- 0.04
	XX 12-707	10.0 mg/kg	5M/5F	1.00 +/- 0.61	0.80 +/- 0.45	0.90 +/- 0.52	0.87 +/- 0.08
	Positive control	30 mg/kg	5M/5F	13.70 +/- 4.96	12.50 +/- 2.78	13.10* +/- 3.84	0.95 +/- 0.07
48 h after treatment	Vehicle control	10 ml/kg	5M/5F	0.80 +/- 0.27	0.90 +/- 0.22	0.85 +/- 0.24	0.92 +/- 0.07
	XX 12-707	10.0 mg/kg	5M/5F	1.10 +/- 0.22	0.70 +/- 0.57	0.90 +/- 0.46	0.77 +/- 0.14
	Positive control	30 mg/kg	--	--	--	--	--

PCE polychromatic erythrocytes

NCE normochromatic erythrocytes

PCE(M) micronucleated polychromatic erythrocytes (in per mill)

NCE(M) micronucleated normochromatic erythrocytes (in per mill)

MV mean value

SD standard deviation

(1) calculated on basis of 1000 NCEs scored per animal

different effects in Male and Female regarding this behaviour, no pooling recommended

* statistically significant as compared to vehicle control, p<0.05

(positive control: statistical analysis was only performed with pooled values for both sexes)

CONCLUSION

Using ODS and proc report effectively we were able to automate a previous manual and cumbersome report production. SAS Stored Processes give us the means to empower the scientists to start a quite complex statistical analysis on their own; the different output files still facilitate the in depth analysis and approval of a statistician.

REFERENCES

US FDA/CFSAN. 2000. "Toxicological Principles for the Safety Assessment of Food Ingredients : IV.C.1.d. Mammalian Erythrocyte Micronucleus Test", <www.foodsafety.gov/~redbook/redivc1d.html>
Olinger, C.R., Tobias, R. D..(1998). " ODS for Data Analysis: Output As-You-Like-It in Version 7", Paper 188, NESUG 1998 <<http://www.nesug.org/proceedings/nesug98/stat/p188.pdf>>
SAS Institute Inc. "Suppressing Execution Options for the SAS Stored Process Web Application Property Page", <<http://support.sas.com/techsup/technote/ts764.pdf>>
LaLonde, Steven M.. 2008. "ODS for the Professional Statistician", Paper 377-2008, SAS Global Forum 2008

RECOMMENDED READING

Cooman, Franky De and Veerbeek, Rudy. 2006. "Strategic roadmap for the IT support of the analysis and interpretation of data in drug discovery", Paper TS08, PhUSE 2006

PhUSE 2008

Hall, Angela and Miles, Brian. 2007. "SAS® Information Delivery Portal – A Tutorial", Paper 214-2007, SAS Global Forum 2007

Pope, David (2007). "Empowering Your SAS Business Intelligence End Users via a SAS Stored Process", Paper 038-2007, SAS Global Forum 2007

Pratter, Frederick E. (2007). "Delivering Dynamic Content with the SAS® Stored Process Web Application", Paper 231-31, SUGI 31

Pratter, Frederick E. (2007). "SAS Enterprise Guide and Stored Processes", Paper 218-2007, SAS Global Forum 2007

SAS Institute Inc. 2007. SAS Institute white paper. "SAS® Stored Processes – Introduction and Overview"
<support.sas.com/documentation/whitepaper/downloads/101519_1104.pdf>

CONTACT INFORMATION

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

Volker Harm
Bayer Schering Pharma AG
Müllerstraße 178
13342 Berlin
Work Phone: +49-30-468-11208
Fax: +49-30-468-91208
Email: volker.harm@bayerhealthcare.com
Web: <http://www.bayerscheringpharma.de>

Brand and product names are trademarks of their respective companies.