PhUSE 2007

Paper TS11

Using SAS Stored Processes and the SAS Portal for Delivering Statistics to Drug Discovery

Volker Harm, Bayer Schering Pharma AG, Berlin, Germany

ABSTRACT

Using SAS® Stored Processes and the SAS® Information Delivery Portal our department Global Drug Discovery Statistics is building up a statistics portal delivering SAS® applications to scientists in Global Drug Discovery. We use this to convert newly developed statistical solutions into standard routine methods as well as to provide applications implementing established standard evaluations previously done by our department.

This will be exemplified by a little application that uses control charts to determine control ranges for IC50-values used in cytotoxicology.

Given the infrastructure of SAS® Business Intelligence this is a really effective way to have applications centrally developed and maintained and to deploy them enterprise-wide without local installation efforts serving our mission "Do more statistics!".

INTRODUCTION

DRUG DISCOVERY STATISTICS

In contrast to clinical research, where the statistical evaluation of studies takes place in a regulated environment, in drug discovery no processes exist in drug discovery that require analyses done by a statistician. Scientists who apply statistical methods form small dislocated groups with very special requirements. This makes application development in large scale economically unfeasible.

SUPPORTING DRUG DISCOVERY

So how does our department drug discovery? Naturally, at first we had to build up a reputation on applying statistics to research problems. This can be done by giving presentations, teaching and consulting. At some point you get involved in developing new methods in research which require specific statistical methods, or you find used methods that definitely have to be enhanced. Now statistical advice is given, if the problem is known, or the literature is searched for adequate statistical handling. A statistical report is written to document the results.

And the statistician or an statistical analyst applies this new knowledge to the first data at hand by programming a typical one-off SAS program. This is done for a couple of times until there is confidence in the structure of data and adequacy of the method. This is the time when data structures can be defined and macros can be written: A standard routine has been established.

THE STATISTICS PORTAL

When our company migrated to SAS®9, we saw a chance to even a step further. As we now had access to an installation of SAS® Enterprise BI Server, we started testing SAS® Enterprise Guide, SAS® Stored Processes, the SAS® Stored Process Application and the SAS Information Delivery Portal. The first results looked promising and so we decided to go a step further and develop some prototype application, which can show the feasibility of a Statistics Portal as a means to deliver statistics to the scientists. Given the wealth of standard routines at hand, which can easily be adapted, centrally deployed and managed, this truly could be a means to deliver statistics to the scientists.

SAS® BI INFRASTRUCTURE USED

We tried to use the SAS BI Infrastructure as straightforward as possible and used the components as provided by the standard installation. The SAS Enterprise Guide was used to create the stored processes. Only basic functionality of the SAS Information Delivery Portal providing links, collections and the ability to execute stored processes was sufficient to build up the web pages of our prototype portal.

STATISTICAL SUPPORT – THE TRADITIONAL WAY

In the case given we were asked to give advice, if there is a better way to assess the validity of experiments in cytotoxicology.

THE STATISTICAL REPORT

The problem was analyzed and a report was written. The report states in its introduction:

"As part of cytotoxicology experiments IC50 values of different compounds are determined. In addition to different compounds, the IC50 value of Menadione (positive control) is determined per experiment.

Current practice is that the validity of the experiment is assessed by comparing the IC50 value of Menadione to the mean ±2std of historical values of IC50 values of Menadione. Values that are outside the boundary indicate a possible validity issue.

This report introduces control charts for this process, and will recommend some small changes to the current practice. Furthermore, control charts are given for selected parameters."

And the solution is described as:

"We will create a control chart for individual measurements and moving ranges. These are special forms of so-called "Shewhart charts", that are used in process control. The respective IC50 values are log-transformed to approach a normal distribution of the measurements.

Two graphs are provided. One figure contains the individual measurement chart (I-chart), in which the individual (log-transformed) IC50 values and the control ranges (control limits) are displayed. The moving range chart (R-chart) contains the ranges of the two most recent measurements. Also in this plot control ranges are displayed."

THE SAS PROGRAM

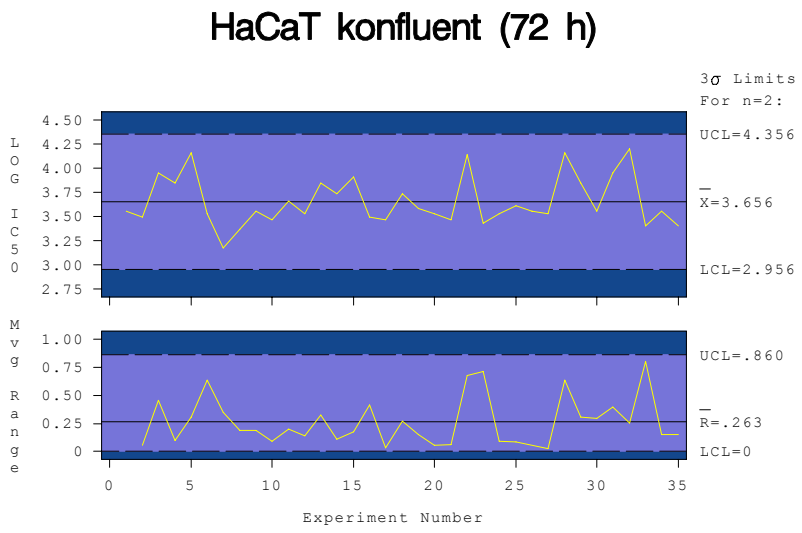
Sounds complicated, or? Fortunately in SAS®/Stat there is proc shewhart, and proc shewhart is capable of producing exactly the proposed control charts. Hence as the result of the analysis we not only have a statistical reasoning for a new method, but also a quite simple way to produce the control charts. A one-off SAS program that produces the control charts is provided in an appendix to the report:

```
libname data 'c:\work\projects\cytotox';
%macro shewhart(data, graph);
    data tt;
        set &data;
        lic50 = log (ic50);
        label nr = "Experiment Number" lic50 = "LOG IC50";
    run;
    proc shewhart data = tt;
          irchart lic50*nr / cframe = vigb
                             limitn = 2
                             cconnect = yellow
                             cinfill  = vlib;
    run;
%mend shewhart;
title1 "HaCaT konfluent (72 h)";
%shewhart(data.hacat_k72, hacat_k72);
```

The data to be evaluated were collected in an Excel sheet containing the data for the cell line (here HaCaT konfluent). For this first evaluation the SAS data sets were created with cards statements.

The macro shewhart transforms the IC50 values and calls proc shewhart. Proc shewhart produces the above mentioned individual measurements and moving range charts.

And this is the graph the program produces:



THE STANDARD PROCEDURE

As this is easily done as a one time task, in a routine process it is a bit more tedious: There are four cell lines, for

which these values are determined. Each time an experiment is conducted, new values will be added and new control ranges should be determined. In the traditional way that would mean for the scientist: Deliver the new data to the Global Drug Discovery Statistics Department and wait for the evaluation. And for us: Add the new data, run the program, and deliver the results to the scientists.

EXTENDING STATISTICAL SUPPORT – THE SAS PORTAL SOLUTION

STORED PROCESS "NEW IC50"

While in the traditional way the main focus lies on the number crunching, in providing an application used by the scientist, we also have to consider the requirements on the data from the scientists view. Therefore the first stored process we have to develop is one to get new data not the highly sophisticated evaluation. This needs a bit of requirements analysis together with the user. Whereas in our one-off program only the statistically relevant variables IC50 for the IC50 value and nr for the number of the value appeared, our stored process for adding values to the database has to handle all the parameters relevant to the scientist.

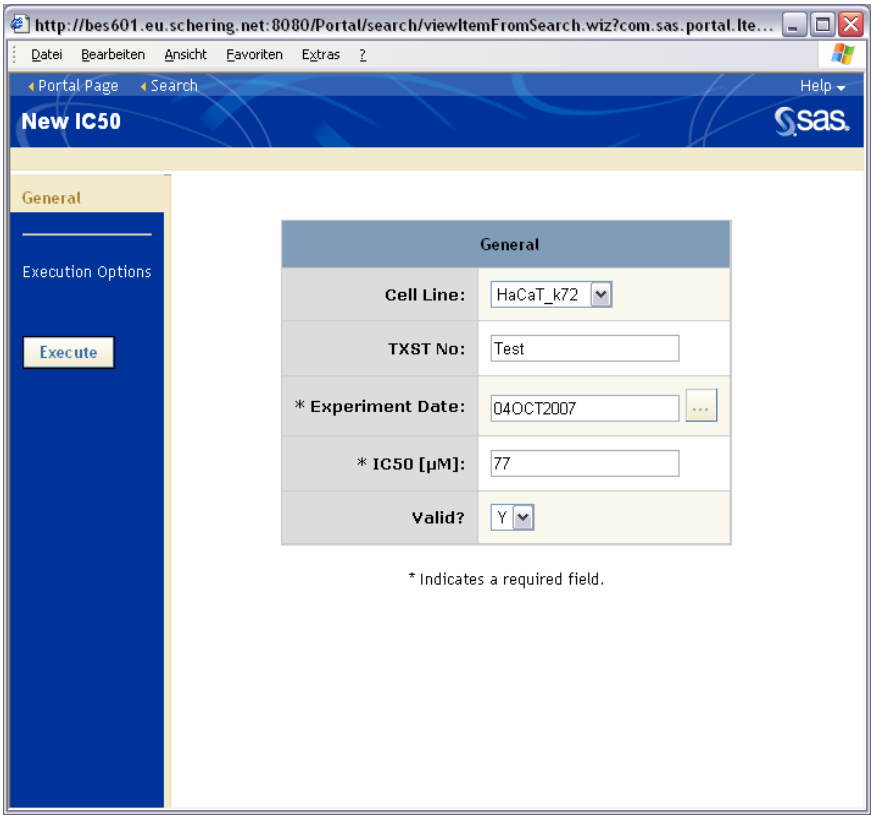
The code of the stored process created with SAS Enterprise Guide is as follows:

```
*  Begin EG generated code (do not edit this line);
*
*  Stored process registered by
*  Enterprise Guide Stored Process Manager v4.1
*
*  ================================================================
*  Stored process name: New IC50
*  ================================================================
*
*  Stored process parameter dictionary:
*
*  _________________________________
*  CELLLINE
*       Type: String
*      Group: > General
*      Label: Cell Line
*       Attr: Visible, Modifiable
*    Default: HaCaT_k72
*  _________________________________
*  EXPERIMENTDATE
*       Type: Date
*      Group: > General
*      Label: Experiment Date
*       Attr: Visible, Modifiable, Required
*  _________________________________
*  IC50
*       Type: Float
*      Group: > General
*      Label: IC50 [µM]
*       Attr: Visible, Modifiable, Required
*  _________________________________
*  TXSTNO1
*       Type: String
*      Group: > General
*      Label: TXST No
*       Attr: Visible, Modifiable
*  _________________________________
*  VALID
*       Type: String
*      Group: > General
*      Label: Valid?
*       Attr: Visible, Modifiable
*    Default: Y
*  _________________________________
*;
*ProcessBody;

%global CELLLINE
        EXPERIMENTDATE
        IC50
        TXSTNO1
        VALID;

%STPBEGIN;
*  End EG generated code (do not edit this line);
```

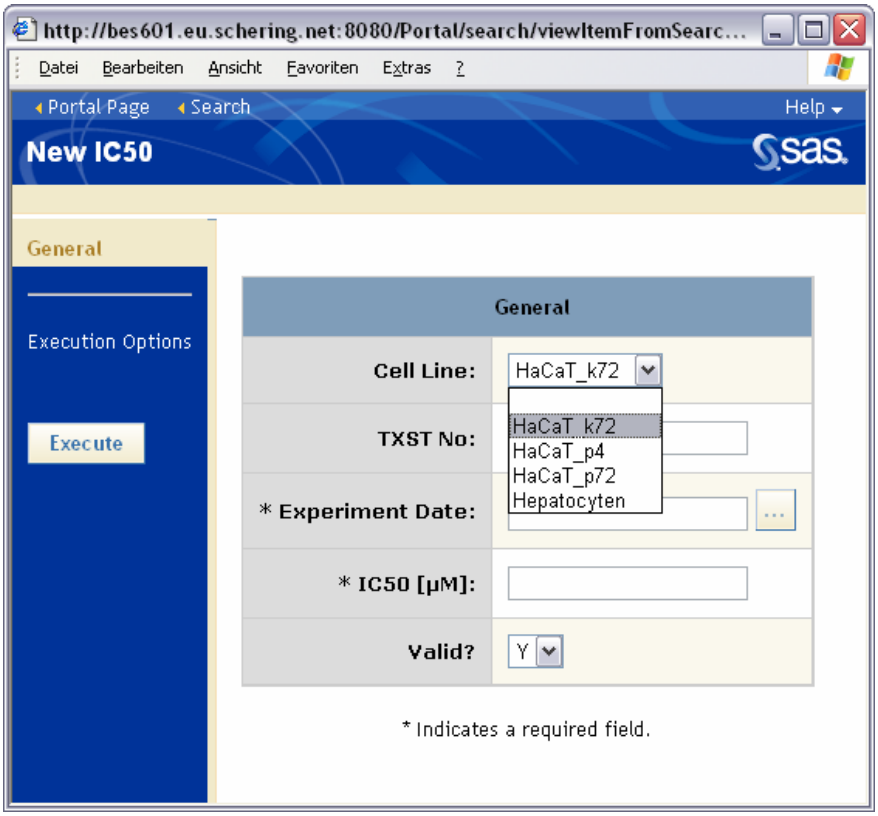
```
%let StoredProcessName = New IC50;

* name and version;
%include "D:\CompStat\dev\Cytotoxicology\QCCytotox\Demo\SASMacro\NameVersion.sas";
* cell line types;
%include "D:\CompStat\dev\Cytotoxicology\QCCytotox\1.0\SASMacro\CellLineTypes.sas";
* repository settings;
%include
"D:\CompStat\dev\Cytotoxicology\QCCytotox\Demo\SASMacro\RepositorySettings.sas";

%macro GetData;
    data wk.New;
        length
            CellLine         $ 11
            TXSTNo1          $ 30
            ExperimentDate     8
            IC50               8
            logIC50            8
            Valid            $  1;
        format
            ExperimentDate IS8601DA10.;
        label
            CellLine = 'Cell Line'
            TXSTNo1 = 'TXST No'
            ExperimentDate = 'Experiment Date'
            IC50 = 'IC50 [µM]'
            logIC50 = 'logIC50 [µM] (derived)'
            Valid = 'Valid';
        CellLine = "&CellLine.";
        TXSTNo1 = "&TXSTNo1.";
        ExperimentDate = input ("&ExperimentDate.", date9.);
        IC50 = input ("&IC50.", best8.);
        logIC50 = log (IC50);
        Valid = "&Valid.";
        output;
    run;
%mend;
%macro UpdateDB;
    data &HistIC50MenadioneDB.;
          set &HistIC50MenadioneDB. wk.New;
    run;
%mend;
%macro ReportResult;
    title "&NameVersion.: &StoredProcessName.";
    title4 "The following observation was added to the database for cell line
&CellLine.:";
    proc print data = wk.new;
    run;
%mend;
%GetData;
%UpdateDB;
%ReportResult;

*  Begin EG generated code (do not edit this line);
*';*";*/;run;
%STPEND;

*  End EG generated code (do not edit this line);
```

Let us see, what is in there: In the first section “Stored Process Parameter Dictionary” all parameters and their metadata are listed. The actual code starts with the global macro variables according to the parameters. The macro GetData evaluates the parameter values and creates a data set with one observation. Macro UpdateDB appends this data set to the original one. Finally the added observation is sent to the output window.
SAS Enterprise Guide creates an input form with the following nice features:

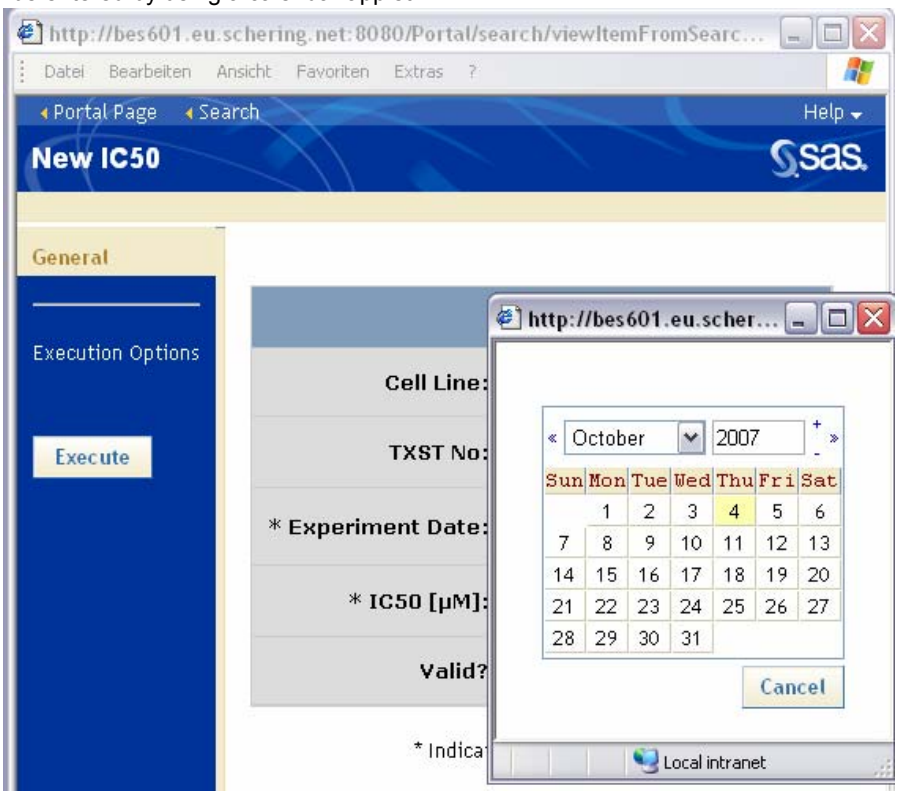
- Required fields are marked with an asterisk.



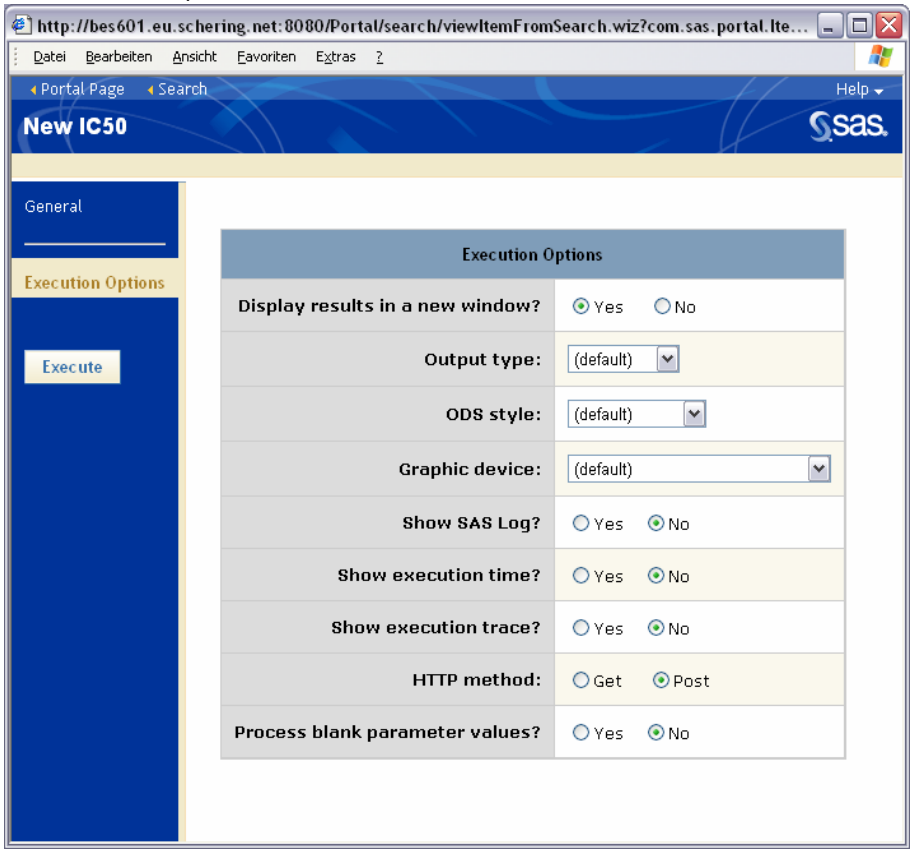
- For Parameter CellLine a list of valid values is used.



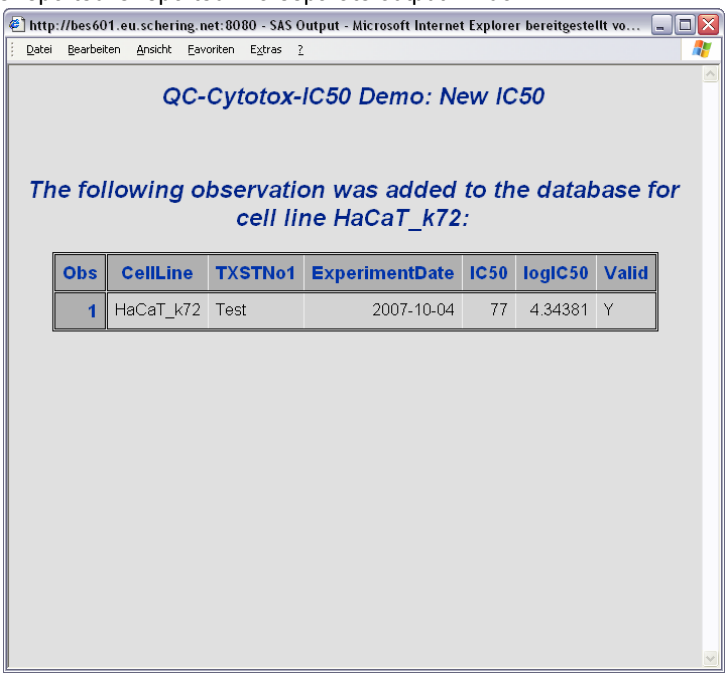
- Dates can be entered by using a calendar applet.



- Standard SAS execution options can be selected.

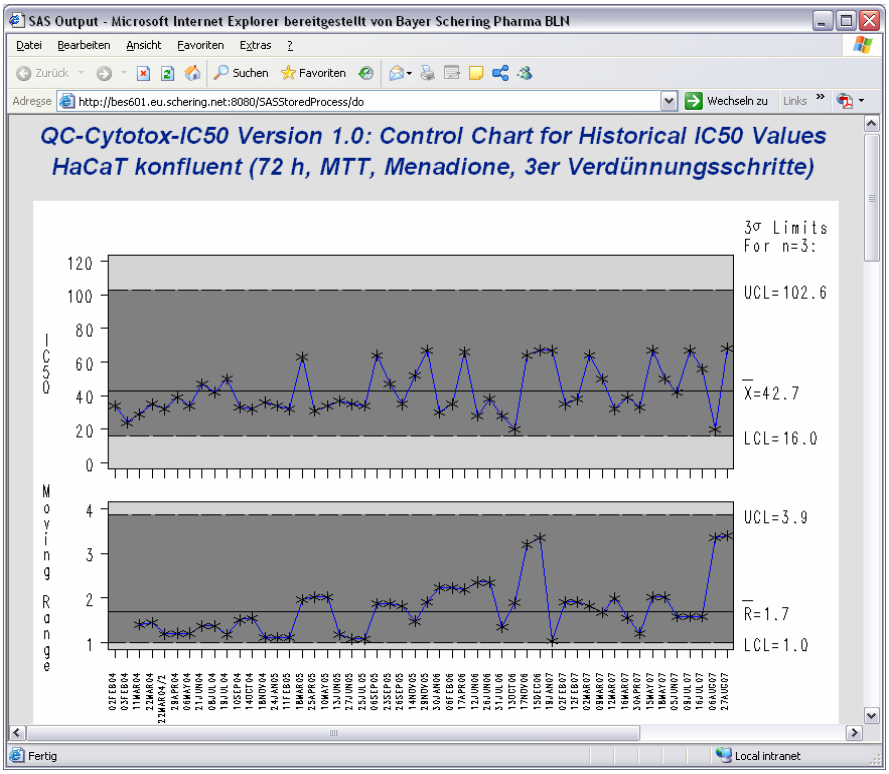


- The added entry is reported is reported in a separate output window.



STORED PROCESSES "DISPLAY CELL LINE" AND "DISPLAY CONTROL CHARTS"

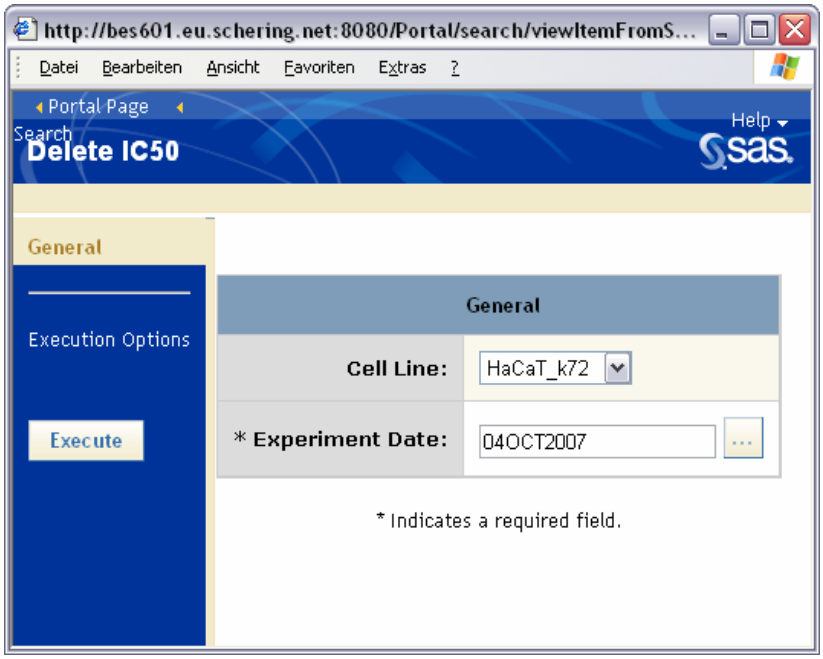
Control charts are created by the stored processes "Display Cell Line" and "Display Control Charts". "Display Cell Line" selects data of one cell line from the database, calculates the control values from the last 50 valid entries and displays the control charts as well as the data of the cell line. Following figure shows partial output of "Display Cell Line". Note that there now dates as tick marks on the horizontal axis. For the users the IC50 number was replaced by the experiment date as key.



The stored process "Display Control Charts" calculates the control ranges for all cell lines and displays the control charts without data listings.

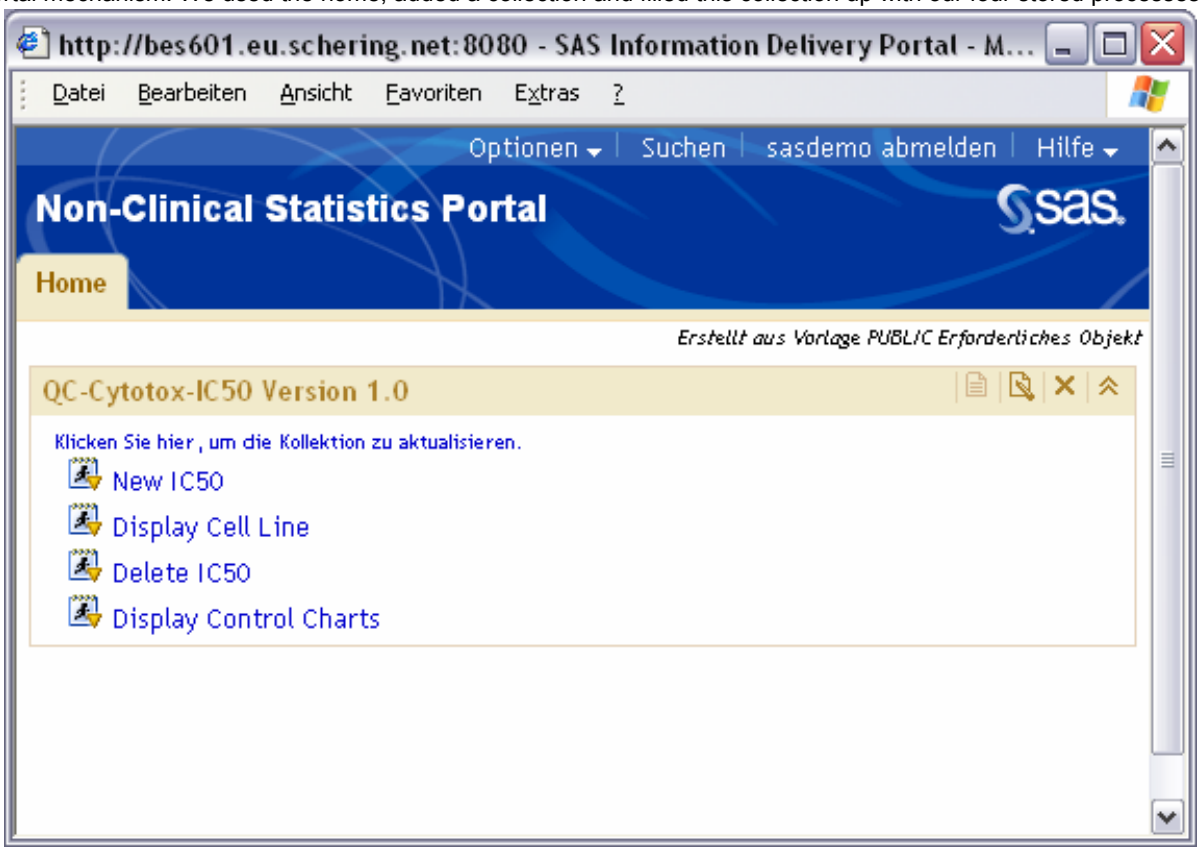
THE STORED PROCESS "DELETE IC50"

Just for completeness we need a stored process "Delete IC50" in case mistakes are made during data entry. As pointed out before the IC50s are identified by the experiment date.



THE SAS PORTAL

With the four above described stored processes we now have complete application for determining the validity of these cytotoxicology experiments. What needs to be done is to put them into the portal. This is done by standard portal mechanism. We used the home, added a collection and filled this collection up with our four stored processes.



CONCLUSION

I think stored processes really a new dimension to statistical support. Of course it more a matter of presentation than of statistics. But if we want the scientist to "Do more statistics!", that is what we have to do.

RECOMMENDED READING

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

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