Paper 020-2008

The Design, Development and Implementation Of A Control Data Tracking And Analysis System For Immunoassays Using SAS/AF

How Tsao, PhD, Associate Director, EPP, Wyeth
Andrew McKeen, Senior Principal Biostatistician, Preclinical Biostatistics, Wyeth
Roger French, PhD, Director, Preclinical Biostatistics, Wyeth
Andrew Bascom, Associate Director, EPP, Wyeth
Joseph Kuo, Research Scientist II, EPP, Wyeth
Khushroo Shroff, Senior Director, EPP, Wyeth

Abstract

The Early Phase Programs (EPP) group at Wyeth Vaccines Research carries out immunoassay development, qualification, validation, and high throughput testing to support pre-clinical vaccine studies and clinical trials. The tracking and analysis of all data collected from control samples that accompanies assay runs is a key function of the throughput operations. It allows routine surveillance of testing processes and assay performance and enables the laboratories to detect and investigate sources of process variability. However, tracking control data in real time is difficult without the proper tools. A multidisciplinary team designed and implemented a new system to provide real time control data tracking. This system uses a SAS/AF interface integrated with other laboratory information systems, forming the main engine for data extraction, analysis, and reporting.

This system uses the flexibility of SAS/AF in combination with SAS/GRAPH, including its Annotate Facility, SAS Macro Language, SAS/ACCESS for Oracle, PROC SQL to create tables, views, and macro variables, ODS, including the TEMPLATE Procedure, and the DATA step to create a dynamic set of on-screen results and reports in PDF format. This presentation will also describe the strategy, process definition, system architecture, and the methods developed to manage the complexity and implementation of the system.

Key Words

SAS, SAS/AF, Control Data Tracking, Assay Throughput, System Integration, System Engineering

Introduction

Wyeth is in the business of making novel vaccines that help prevent infectious diseases. One of our vaccine R&D efforts focuses on producing vaccines against pneumococcal infection. The EPP group in Wyeth is responsible for assessing immune responses to vaccines in human trials. High throughput testing of clinical samples is carried out under regulated conditions. All immuno-assays that are run in high throughput format have control serum samples; the data from these provide a useful measure for the quality of assays over time. When the testing environment is unchanged, a control data series is expected to display a steady, consistent, and random scatter around a center with no patterns or trends. Control data tracking and analysis is a key activity in any testing laboratory; it provides routine surveillance of the performance of testing processes and enables the laboratories to drive decisions and investigate sources of process variability.

In this article, we share our strategy for improving the control data tracking and analysis process using SAS. We will discuss the processes and methods that were applied as well as issues that were encountered and resolved.

Case for Action and Objectives

A continuous process improvement initiative by the business operations manager in 2005 required that an integrated system be built to track, analyze, and report control data results. The new system, named Control Data Tracking and Analysis System (CDTAS), was to replace legacy systems and achieve the following objectives:

- Visibility of control data by management and operators – people need to access and "see" the data anywhere, anytime.
- User-friendly interface and timely feedback of newly entered information
- Standardized procedures across key assay platforms enabling integration of new assays into the system without significantly changing the input or reporting structure
- Alignment with the laboratory information management system architecture, LabWare LIMS
- Availability of a rich collection of credible and reliable statistical procedures - to meet a growing need in statistical graphics, data analysis, quality control, and reporting
- Inclusion of critical process information relevant to potential sources of variability
- Integration with daily assay process workflows

Building CDTAS Using SAS Solutions

SAS/AF was chosen for the development of CDTAS interface because SAS is the preferred data analysis tool for both the Preclinical Biostatistics and the Early Phase Programs group at Wyeth. The underlying tools needed are all available in SAS and system maintenance and support could be supplied from in-house and consulting resources.

Introduction to System Design

The CDTAS system has three main components – the LIMS database, the CDTAS database, and the SAS/AF reporting modules. The LIMS database provides an interface for laboratory personnel to capture daily control data results. This feeds data into the CDTAS database, which serves as a secure, central warehouse for all control data, formatted for statistical analysis. Flexibility was an important consideration in its design so that, over time it could be expanded to include data from new assays and new assay types.

Currently, there are twenty-five tables in the CDTAS Oracle relational database. These tables were designed to capture important results from control samples, but also to facilitate change management - without the need to modify the existing SAS program. Changes include assay platforms, control data parameters or specifications; activation or deactivation of control samples; process related events, new instruments, data fields or formats; and types of control samples, studies or user roles.

The database is also central to the CDTAS system itself in the sense that it forms a fixed point between the changing data acquisition processes in the laboratories and the developing SAS analysis and reporting modules.

CDTAS uses SAS/AF to bridge the gap between the Oracle database and a user-friendly application. The interface allows the users to inspect and summarize data on an ongoing basis both interactively and in report format. This is achieved by a set of SCL commands in combination with modules of SAS code making use of the extensive capabilities of BASE SAS, SAS/ACCESS for Oracle, SAS/GRAPH, and SAS/QC.

SAS/AF Interface Design Considerations

Laboratory personnel responsible for monitoring and improving assay performances routinely view, graph, and generate reports of control data. They must be able to select the correct data, summarize and display it in a way that is useful, and save the results for future reference. A structured interface was designed to mediate the process of data extraction with the option to create plots, charts, and reports.

The interface's SCL code manages and calls SAS modules via Submit Blocks, executing both internal code and external macros. It uses the SAS/ACCESS to access the Oracle database.

Since the tasks to be performed occur both on an ongoing basis and periodically, it was necessary to design an interface able to handle both types of tasks. To keep the number of frames needed to a minimum, we used the Tabular object (a SAS/AF version 6 object) on the main frame with one tab for each type of task. These were labeled Quick Statistics and Summary Report, respectively.

The Tabular object allows the user to toggle back and forth between tasks without losing work performed on the previous tab. This does, however, require the program to keep careful track of data set, macro variable, and SCL object names associated with each task. Good practice in any programming project, but imperative in our application since the users can be, in a sense, running two analyses simultaneously.

The Quick Statistics Tab enables a variety of interactions with the control data to allow problems to be identified and explored. Plots are actively linked to data tables facilitating this process. The Summary Report Tab constructs formal reports for periodic review.

Quick Statistics Tab

Selecting the Quick Statistics Tab (refer to Exhibit 1) displays a panel of three sub-tabs – Data Selection, Data Display, and Graphical Display. Originally, the data listing and its associated plot used the same screen real estate. However, the use of sub-tabs allowed us to expand the amount of screen real estate available for displaying the data and reduce the crowding that was occurring by trying to place a number of AF objects on the same screen.

Exhibit 1 – Quick Statistics Data Selection

The Data Selection tab displays a series of drop down list boxes, radio boxes, and table viewers to allow the user to select a certain subset of data. The sequence of choices is structured to facilitate selection of commonly requested subsets. An external SAS module populates the AF objects with data sequentially, using PROC SQL to create datasets, macro variables, and views. This allows the user to fine-tune the data to be inspected.

Exhibit 2: Quick Statistics Data Display

Once the user completes the required selections, the final data set is displayed in a Table Viewer Control on the Data Display tab (refer to Exhibit 2, above), where tables can be re-sorted by selecting any column header.

Please note that the data displayed in Exhibit 2 as well as subsequent graphs in Exhibits 2 ~ 6 were "simulated" to help illustrate different functions of CDTAS; they do not represent real business information at the laboratory.
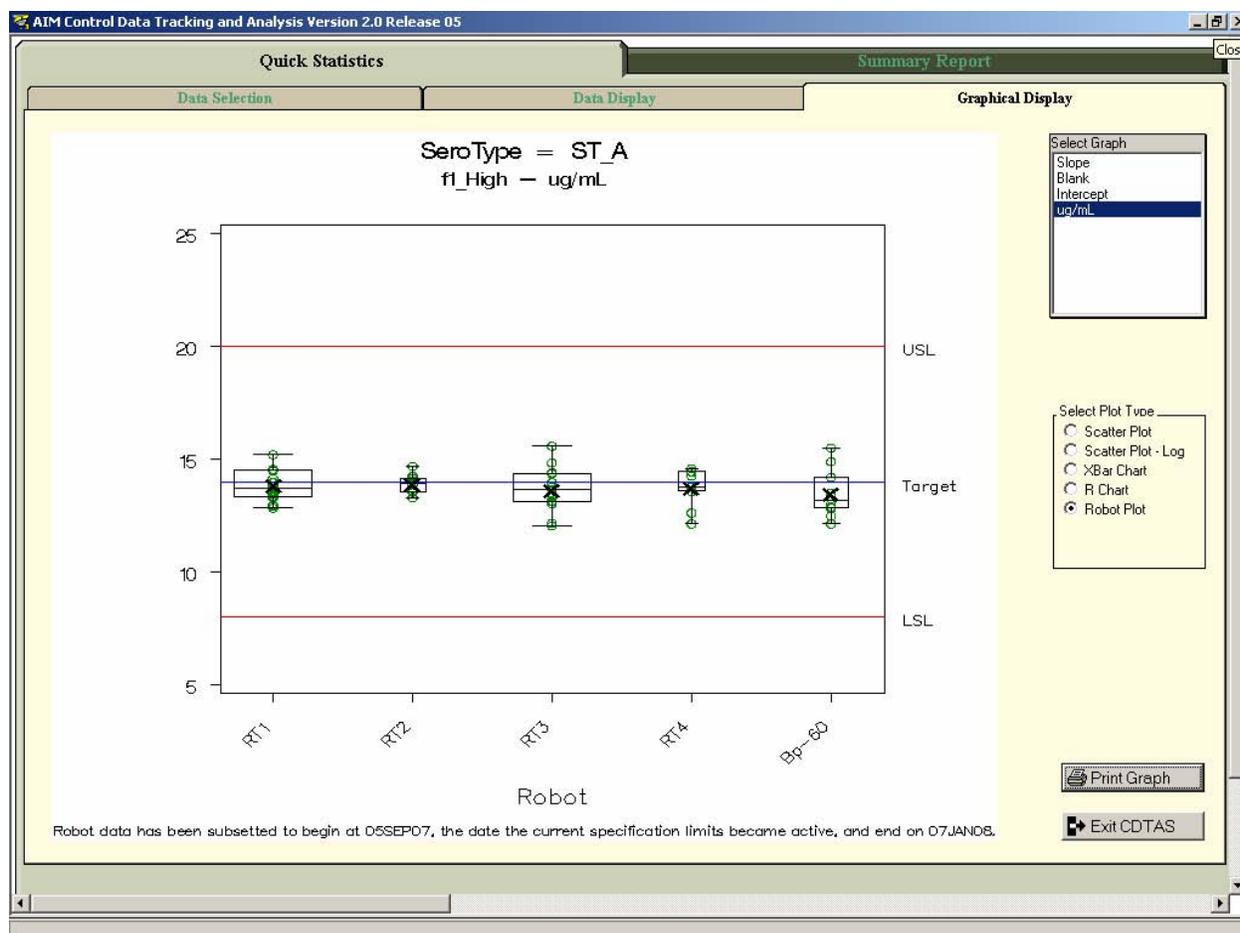
Exhibit 3: Quick Statistics Graphical Display

By switching to the Graphical Display tab (Exhibit 3, above), the user can graph assay data focusing on assay performance over time -comparing control data with specification limits. Types of graphs available to the users are scatter plots, line plots, histograms, box plots, and several Shewhart type control charts.

For the scatter plots, a standard PROC GPLOT graph could have been used. However, we wanted to be able to select values on the graph and have the corresponding records highlighted on the listing. Instead we combined PROC GPLOT with the Annotate Facility of SAS/GRAPH to generate plots, making it easy to determine the SPOTTYPE and SPOTID for data points chosen by clicking on the displayed plot. We use this information to link and highlight the corresponding records in the data table.

We chose this method of linking instead of pop-ups since the number of columns displayed in the data table can be numerous. Note that this link between the plot and table also works in reverse – the user can select multiple records in the data table and the corresponding points are highlighted in the graph. This facilitates the investigation of identified problems.

In addition to the interactivity between the Table Viewer Control and the Graph Output Control objects, the Quick Statistics tab also has several pushbutton controls that allow the users to print the currently displayed data (with any selected data values highlighted) in PDF format via ODS PDF. They may also save the currently displayed data in JMP, SAS7BDAT, and XLS formats, or print the currently displayed graph in PDF format via ODS PDF.

Summary Report Tab

Formal reports are created using the Summary Report tab. Like the Quick Statistics Tab, the Summary Report Tab (refer to Exhibit 4) employs a structured series of drop down list boxes, radio boxes, and table viewers allowing the user to select a subset of the data. The contents of these fields are populated sequentially as the users makes choices using PROC SQL to create datasets, macro variables, and views enabling the user to fine-tune the data to be included in reports. Once the user completes the required selections, data are displayed in a table viewer with capabilities similar to that on the Quick Statistics tab. The user can also save or print the displayed data.

Exhibit 4 – Summary Report

Reports are generated to summarize data using the experimental ODS LAYOUT and ODS REGION statements in conjunction with ODS PDF. General formatting of the report such as page margins, font styles, and font sizes is accomplished by using the TEMPLATE procedures of SAS/BASE.

Each report includes up to three distinct sections: tables, graphs, and Pareto charts; their style and content depend on the type of assay and data chosen. Tables are created within the reports by PROC REPORT of SAS/BASE or by an annotate dataset with PROC GANNO of SAS/GRAPH. These are relatively simple to create. However, the graphics needed in the reports are more specialized.

In order to add more detail about assay performance, all graphs are augmented by extensive use of the Annotate Facility of SAS/GRAPH in conjunction with PROC GPLOT. Refer to Exhibits 5 and 6 for sample graphical output generated by the preceding procedure.

When there is a trend, shift, or change of variability in the control data series, it is helpful to have historical process information at hand. Incidents of key process changes were routinely captured and were displayed as vertical lines in the graph corresponding to the time of occurrences (See simulated example in Exhibit 5). Note that each incident is matched with a temporary incident code with the corresponding incident category information displayed at bottom of each plot.
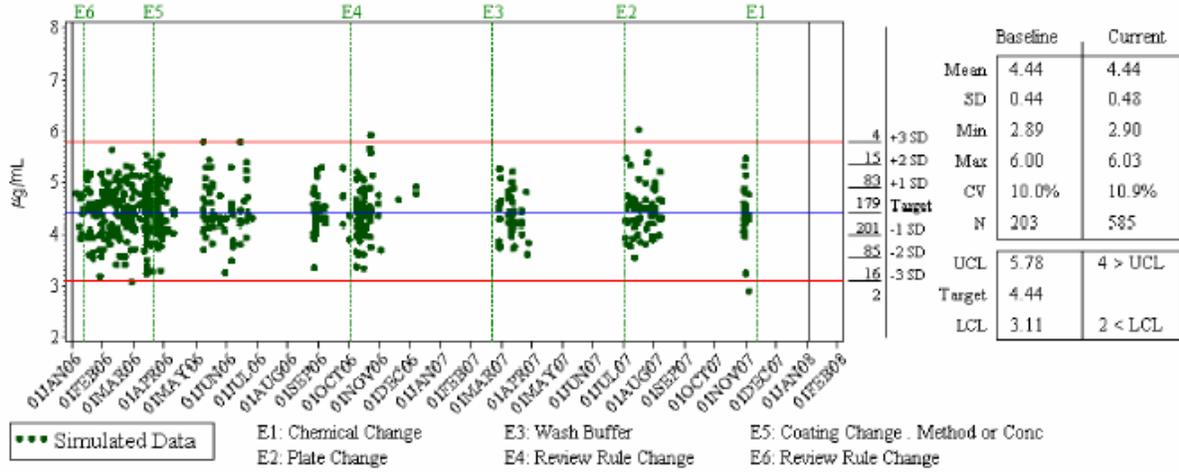

Exhibit 5 – Sample Specialized Graph

One of the challenges we had was to construct a meaningful chart that would display data series with a pre-determined number of possible values; it is hard to see the true frequency of these data points. To help visualize the data, a histogram is displayed to the left side of observed values; the target value, as well as lower and upper specification limits were highlighted in the graph (See Exhibit 6).
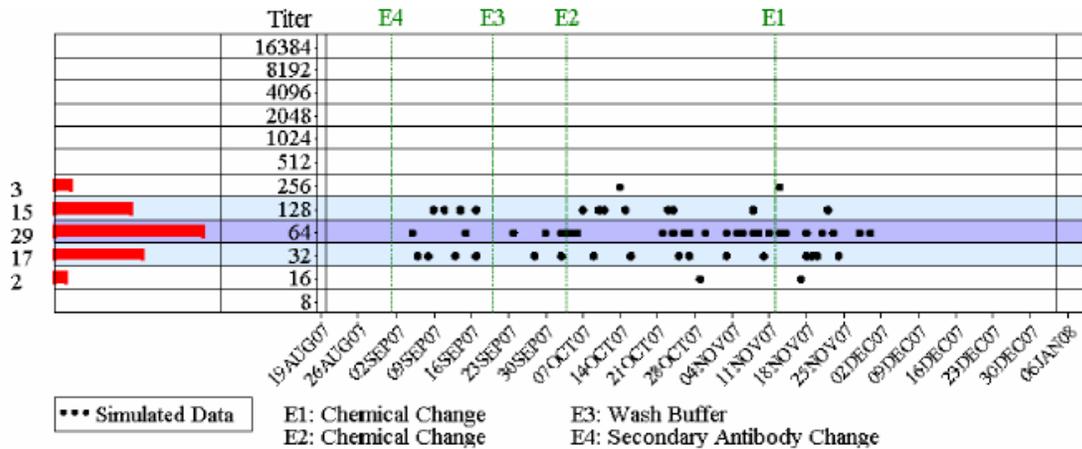

Exhibit 6 – Sample Specialized Graph

Once the graphs have been generated, the report is supplemented with a Pareto chart generated by PROC PARETO of SAS/QC to summarize reasons for rejection of plates.

Making It Work – System Engineering

The principles we applied to build the system were based on Six Sigma's Define, Measure, Analyze, Implement and Control (DMAIC) cycle. We used a phased approach to manage constantly changing priorities, demands, resources, and technologies with a focus on four main threads: Business Process Architecture, Technical Infrastructure, Complexity Management, and Change Management.

The system went live in June 2006 and the subsequent stabilization period has been very smooth. Very few system issues were reported and the problems, when reported, were very minor.

Looking back, there were a few key success factors worth of mentioning. First, a detailed project plan - including a work breakdown structure and all required deliverables - was developed and updated to guide all phases of the project activities. Second, regular gap analysis, requirements and design reviews, and super user participation with system testing ensured the quality of product would meet customers' expectations. Third, developing staffing plans and establishing standard-operating procedures with the business groups integrated the system into ongoing business practices. Fourth, a solid technical infrastructure – including a relational database as the main engine and an input data module for lab personnel – facilitated the interface between SAS and other laboratory information systems; it also allowed the system to manage changes without significant system modification. Finally, great teamwork and strong support by the sponsors of this project were critical.

A Path Forward

Since the system went live in June 2006, it has accomplished most of its initially stated objectives and is now being scaled up to include more assays and more functions. Currently, the system is under Life Cycle Management with two primary assays. Two regular releases have been completed to meet growing business needs.

We now have a growing number of users and a rapidly increasing volume of data accumulated in the system over time. Our primary challenges are to provide access to the system via a web interface while maintaining good processing speed. Also, introducing additional untapped capabilities of the SAS system will help us to better meet continued demand for more effective management, analysis, and reporting of control data.

Acknowledgements

References

SAS Version 9.1.3 WIN_PRO platform, SAS Institute Inc. Cary, NC, USA