SAS/PH-Clinical® Software: CANDA Implementation Strategies
J. Meimei Ma, Quintiles Inc.
Sandra D. Schlotzhauer, SAS Institute Inc.

Abstract
SAS/PH-Clinical® software is currently used for a variety of clinical data review needs within the pharmaceutical and biotechnology industries. Certainly, one of the most critical needs for effective delivery of clinical information comes when implementing a Computer-assisted New Drug Application (CANDA). SAS/PH-Clinical software is an obvious choice to provide the data function of a CANDA system.

The intent of this paper is to provide tips and techniques for utilizing SAS/PH-Clinical software for a CANDA situation. Strategies for developing a successful CANDA system based on SAS/PH-Clinical software are discussed. Emphasis is placed on the significance of effective coordination with the traditional in-house review effort and the benefits of applying that knowledge when setting up a computerized system for regulatory agency use. Given the importance of planning, future enhancements to SAS/PH-Clinical software that would benefit CANDA systems are highlighted. An overview of hardware considerations is included.

Introduction
Since this paper focuses on the use of SAS/PH-Clinical software in CANDA systems, a number of assumptions have been made about the reader. We trust you are familiar with clinical trials research, and in particular with aspects of computerized data management and/or statistical programming. Basic knowledge of current FDA Guidelines is assumed, but detailed knowledge of FDA preferences about CANDAs is not. Being a user of the SAS® System, as opposed to a programmer, is sufficient technical background. Certain comments about using SAS/PH-Clinical software may make more sense if you have seen at least a brief demonstration.

CANDAs and Drug Development
Regardless of therapeutic area, the clinical trials part of the drug development process has standard stages. Figure 1 shows a simplified timeline for the major phases leading up to an NDA submission by a sponsor. Note that activities leading up to the submission may be handled entirely by the sponsor company or could involve a clinical contract research organization (CRO). For the purposes of this paper, we will focus on NDAs although the process is similar for a PLA submission or a CNS submission in Canada.

For the clinical section, final NDA preparation combines data and results from individual studies into integrated summaries of safety and efficacy (ISS and ISE). One way to describe this process is:

Pivotal + Supportive → Integrated Summaries.

In other words, the integrated summaries are composed of the combined information from the pivotal and supportive studies. Therefore, information in a CANDA often focuses on these core studies and integrated summaries.

Since 1984, companies have tried differing approaches in order to supplement the paper NDA with a computerized system. An optimally designed electronic submission provides a reviewer (usually focused on the medical reviewer) access to important, relevant, and helpful information via a user-friendly interface. The primary functions of a "full-featured" CANDA are:

- Data (clinical and/or statistical analysis)
- Text file retrieval (view, cut/paste)
- Indexed image access (CRF and/or text).

The FDA commitment to electronic submissions is explicit and supported directly by the Commissioner's...
Office. After much anticipation, the first version of the CANDA Guidance Manual was released in October 1992 and discussed at a joint FDA/PMIA meeting. A great deal of effort and thought has gone into the production of a useful "living document" for focused development of future CANDA systems.

SAS/PH-Clinical Software Environment

To discuss using SAS/PH-Clinical software in CANADA applications, we must first cover three key concepts: Product Administrators, study definitions, and the general computer environment.

The Product Administrator, or PA, is a special SAS/PH-Clinical user, much like a database administrator is a special user of a DBMS. The PA defines trials, users, and devices, and performs installation and maintenance tasks. For SAS/PH-Clinical software users, a dedicated PA would handle each drug program. This might be accomplished by having a single person handle several drugs. The PA would be working in a multi-user environment. Beyond the PA, other programmers would provide a high level of expertise in using the SAS System for data management and analysis of clinical trials data.

The most important tasks PAs perform involve creating study definitions. Defining a study (clinical trial) to SAS/PH-Clinical software includes identifying the data libraries, defining how often each variable is measured for each subject in the trial, defining variables as appropriate for analysis, and so on. By defining studies, the PA frees all other SAS/PH-Clinical users from these tasks and helps control the accuracy of information provided. After defining a study, users can create patient groups that are composed of some or all patients in the study and some or all variables in the study.

Computer environments vary a great deal in the pharmaceutical and biotechnology industries. The environment of interest for this paper assumes the existence of a centralized coordinating center for data management and programming activities. This center may be part of the sponsor company or a CRO. Hardware (CPUs, printers, workstations or terminals) would be sufficient to handle the database creation and analyses for individual studies and integrated summaries.

General Strategy Tips

As with any computer system, planning ahead is the key to success. The myriad of details that must be accounted for can be mind-boggling without a comprehensive, but flexible, plan. The following presents a few ideas to consider early in any CANADA development process.

Integrated Team Approach

The first step in designing an effective CANADA system is to assemble a small multi-disciplinary team. The technical programming expertise provided by Information Systems (IS) and/or Statistical Programming staff need be integrated with knowledge from Regulatory, Clinical Science (also called Medical Writing), and Biostatistics. However, the most important input comes from FDA representatives. At conferences and workshops in 1992, all FDA speakers stressed the importance and benefits of early, informal, and continual interaction with Sponsors during the CANADA development process. The team should include members from all "companies", i.e., the sponsor, the Agency, and any CRO(s) involved.

In order to facilitate team interaction, focused communication is a key factor. FDA input from the medical reviewer, the statistician, or the chemist may be best obtained by funnelling all questions through the Consumer Safety Officer (CSO). In the same way, a programming team leader should coordinate communications to the programmers. Using this approach, the "core team" can remain small even for large projects. An important interaction on the sponsor's CANADA team should be between Regulatory and Programming. Consider the situation once the CANADA system is installed and the FDA reviewer needs help. Clearly, it is not in the sponsor's interest for the reviewer to be confused about who to call. Note that this level of activity is generally required for development of any CANADA, not just one that uses SAS/PH-Clinical software.
CANDA System Integration

With a team in place, you are ready to determine which CANDA system functions are required. In certain situations, an effective CANDA may only need to include data from key studies and a set of organized text files. Increasing capabilities by adding functionality also increases the complexity of the development and validation process. Interaction with the FDA at this stage leads to the most informed cost/benefit decisions by the sponsor.

Once the core functions are selected, then the job of designing, testing, modifying, and implementing a fully integrated system can begin. Assuming more than one is of interest, features that are based on different core technologies may require programming before they form a logical system. For example, if data and CRF image retrieval are included but are based on software from different vendors, the end-user should be able to switch between them by using a menu instead of less intuitive (typed-in) commands.

Whatever the choice of functionality, the most effective system testing occurs when a CANDA system is used during the preparation of the NDA. Several FDA reviewers feel that any system and/or system feature that proves helpful for NDA preparation is probably also useful for review purposes as well. Even with a thorough design process, modifications are inevitable. The longer the testing or “actual use” phase before delivery to the FDA, the more likelihood that a CANDA system is immediately helpful to a reviewer.

Hardware Considerations

Since sponsors who have submitted CANDAs are aware of the hardware issues at the agency, much has been written or presented on this issue. This section is limited to discussing hardware issues as they relate to using SAS/PH-Clinical software as part of a CANDA.

The interface for SAS/PH-Clinical software is color-coded to help provide ease of learning and ease of use, so a color monitor is vital for a CANDA. In addition, a bit-mapped graphics display device is preferable. Not only does this device provide high-resolution graphics and the ability to access the Explore component (an interface to SAS/INSIGHT® software), it will provide the ability to display optical images. Future releases of SAS/PH-Clinical software may require this type of display device.

For workstations in particular, and for PCs to a lesser extent, tuning of the SAS System is important. For example, performance of SAS software on a VMS workstation can be affected by many of the SYSGEN parameters. As part of CANDA preparation, sponsors should consult the appropriate installation guide to determine initial system parameters.

Using more than one hardware platform may be worth considering. A mainframe, minicomputer, workstation, or microcomputer (PC) may be effective during the development phase. If the reviewer requests a different final platform, then allow time for conversion and testing. To address any host-specific issues, such as running programs in batch, sponsors should consider downloading the CANDA to a workstation for the agency reviewer and downloading to a second workstation to remain at the sponsor site. One advantage of SAS/PH-Clinical software is that the interface operates essentially identically on all hardware platforms.

Data Function

Access to data is one of the core functionalities of a full-featured CANDA. The term “data” is used in the broader sense of any information that reviewers may want to browse, analyze, or review as contrasted to the narrower sense of the statistical analysis data. As with more general issues of CANDA development, careful planning for database design and database preparation is vital. The saved patient group data and ability to create superstudies are useful features of SAS/PH-Clinical software in relation to both development and actual use.

Data Preparation

When planning to provide an FDA reviewer (or any non-programmer) with direct access to a clinical trial database, you should apply integrated database design principles as early as possible. Ideally, CRFs for all studies are designed for clarity and consistency across studies. However, as noted by Sollecito and Ma (1992), starting in the middle of the process is not an
uncommon situation. Data may already exist in inconsistent databases, thus making standardization across studies an important and complex data management task. Although "raw" CRF fields may remain different, analysis variables must have consistently defined values in order to permit useful integrated analyses. Since initial setup to allow access to individual CRF data files is straightforward in SAS/PH-Clinical software, it can be used as a exploratory tool in making standardization decisions. Once an integrated database design exists and the study definition is created for one study, then you can use the FASTDUMP and FASTLOAD utilities to define additional studies rapidly.

Even with fully consistent CRFs across studies, the optimum time to start using SAS/PH-Clinical software is as soon as test data are available. An appropriate time for the PA is to create preliminary study definitions for the CRF information is when initial data are available for ongoing Phase II studies. Note that this does not necessarily imply allowing users unrestricted access to the study data. The security group options should be used to avoid any potential problems of unplanned interim analyses. During Phase III, the final Phase II analysis data sets can be added. By starting early, you have time to consider restructuring analysis data files to optimize performance. Essentially, the Phase II study data forms the basis for a pilot or test of the CANDA system.

Saved Patient Group Data
The patient group is a key concept in SAS/PH-Clinical software. After the PA defines studies to the software, users build patient groups. They choose a study, patients from the study, and variables from the study. Users can include all variables or a subset of variables from the study. They can include all patients, or a subset of patients selected by patient number, or a subset of patients that meet user-defined criteria, or a subset selected both by patient number and by criteria.

Patient groups are divided into two portions: the patient group definition and the patient group data. The patient group definition consists of the SAS code needed to reconstruct the patient group--identifying the study, the variables, and the patients included. The patient group data are generated when the definition is run.

With recent enhancements to SAS/PH-Clinical software, users can save the patient group data as well as the patient group definition. When patient groups are saved, the definition is always saved in the SAS/PH-Clinical Library. The data are saved upon request, and are also saved in the Library. Thus a CANDA system can include both the definition and the data of a "frozen" copy of patient groups. While the saved patient group data require more storage space than the definition alone, the data are immediately accessible to the FDA reviewer. Generally, accessing the saved patient group data is quicker than rebuilding the patient group data from the patient group definition.

When sponsors provide existing patient groups, this does not limit the agency reviewers to only those patient groups. All users can generate new patient groups--specifying their own choices for which patients and variables are to be included. In addition, users can begin with an existing patient group provided with the CANDA, subset to exclude patients, and save the subset as another patient group. For example, a reviewer may look at a patient group composed of all patients with adverse events, subset to look at only patients with AEs involving cardiac or liver problems, and save those patients as a new patient group.

Superstudies
Although both sponsor and agency reviewers are interested in examining data for specific studies, they typically want to view combined data such as integrated safety data. In SAS/PH-Clinical software, combining data across studies is done by defining a superstudy. After the PA has done so, users can create patient groups from the superstudy. Once a patient group is generated, all SAS/PH-Clinical tasks can be performed--whether the patient group is formed from a single study or from a superstudy.

Two important issues for superstudies are patient identifiers and visit identifiers. (The updated SAS/PH-Clinical documentation contains a new chapter

Sponsors who uniquely assign patient numbers across all studies for a drug are in a better position to define superstudies. Typically, data for one study are stored in a group of data sets, and data for a second study are stored in a different group of data sets. When the data sets for the two studies are combined, unique patient identifiers must exist across the two studies. If the patients are numbered uniquely only within each study taken separately, then the superstudy definition must include pre-processing (defining SAS data views, for example).

Another key aspect of defining superstudies is to ensure that visit identifiers make sense in the context of the superstudy. Consider the following situation:

<table>
<thead>
<tr>
<th>Study</th>
<th>Visit</th>
<th>Week</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>8</td>
</tr>
</tbody>
</table>

Patients in the first study are examined every 4 weeks, but patients in the second study are examined every 2 weeks. If the data sets from the two studies are merged by visit, then inappropriate comparisons will be made; since visit 3 is the 8-week visit for study 1 and the 4-week visit for study 2. In this situation, SAS views can be used to perform pre-processing on the data to ensure appropriate merges are performed.

**CANDA System Documentation**

An important, and often neglected, feature of any end-user system is documentation. Technical documentation designed for programming maintenance must be kept separate from user documentation that provides guidance about how to use system features. Basic levels of documentation, of both types, should be created and maintained during the entire development process. Creating a complete set of documentation is then a straightforward process once a stable version of the CANDA exists. Remember that modifications must be allowed since the system is certain to change.

SAS/PH-Clinical software includes several helpful documentation features. These include the:

- Overviews component
- Library component
- Tables component
- Output documentation
- Administrator reports on studies and security groups.

CANDA system users will certainly find reviewing study summaries and regenerating summary tables an absolute necessity. Viewing study summaries, both general information and specific tables, can be performed with the Overviews and Library components. Regenerating many summary tables can be done using the Tables component. For the remaining tables, the Library provides the ability to generate, edit, and store programs to generate the tables as well as the ability to store output or graphs resulting from the programs. For any printed output, the output documentation provides all the SAS code necessary to recreate the patient group and the output.

**Overviews Component**

Sponsor reviewers are familiar with each protocol for a drug and may or may not need the Overviews component to remind them of particular details for a given study. However, agency reviewers are asked to assimilate large amounts of information about many drugs from several sponsors, and are typically reviewing several drugs simultaneously. The Overviews component is designed to quickly provide the important details on a study—online, so reviewers need not search through volumes to find a specific protocol.

In addition, sponsors can identify variables to be displayed in the Overviews component, such as the investigators or sites for trials. In this way, an agency reviewer can scan through all the investigators involved in trials for a drug. Or, by using the Subsetting feature, reviewers can view only information on a specific study, for Phase III trials, for a given investigator, and so on.

By providing summary information for each study and by using meaningful study overview variables, sponsors can enhance agency reviewers abilities to quickly find information on a specific study in the NDA.
Library Component

In addition to saving patient groups in the SAS/PH-Clinical Library, users can save programs, output, graphs, and notes. This provides not only a way to document the work performed but a communication tool.

Sponsor reviewers can save programs generated while working with SAS/PH-Clinical software, save the output or graph created as a result of the program, and save any notes to communicate comments to each other. These activities can occur while the trials are in progress, and while the sponsor is performing efficacy and safety analysis after the trial is completed. They can also be an integral part of a CANDA.

When you save programs, graphs, or output in the Library, you can give names and descriptions to identify how the item is used in the CANDA. The program that generated Table 12A in the NDA may be named TABLE12A and given a description that identifies where the table appears in the NDA—which section or volume, for example. Similarly, the output or graph resulting from the program can be saved with an appropriate name and description.

This process can be used whether or not the program is generated using SAS/PH-Clinical software. By design, the software does not perform complex statistical analyses, leaving that to the expertise of statisticians who are familiar with the intricacies of the protocol and the appropriate analysis methodologies. Much analysis is done using SAS software, however, and the programs and output for this analysis can be stored in the SAS/PH-Clinical Library. If necessary, statistical reviewers at the agency can further modify the program, as described below.

For programs, the Library includes features to include SAS programs from another SAS library or an external file. These programs can include additional LIBNAME statements to identify data sets not allocated when SAS/PH-Clinical software is invoked. Suppose a sponsor wants to include pharmacokinetic analysis as part of the CANDA. A program in the Library can identify the pharmacokinetic data (using LIBNAME statements) and analyze the data using SAS software. (See Klonicki 1991 for an example of using SAS/ETS® software to perform model-independent pharmacokinetic analysis.)

For programs, the Library includes the Program Summary, which identifies the appropriate data for the program, explains when and how to use the program, describes the purpose of the output, identifies the programmers, and provides space for additional comments. By providing this information in the program summary, you can increase the ability of the reviewers to quickly choose which saved program best meets their needs.

For graphs, output, and notes, the recommended approach is for the sponsor PA to add the items to the Library. Graphs can be saved in the Library or can be written to a Graphics Stream File (GSF). As part of developing the NDA, sponsors may want to save graphs in GSFs and then include those graphs in the word-processing document used for the NDA. Saved notes can highlight comments about a particular program, graph, output, or patient group. This can facilitate clinician-to-clinician communication: by using appropriate names and descriptions, agency reviewers can quickly see if the sponsor reviewer had any special comments about a graph or table.

By using subsetting in the Library, reviewers (at either the sponsor or agency) can find all items associated with a specific patient group or study. If, for example, an agency reviewer re-analyzes one table for a patient group and wants to find all other tables based on that patient group, the Subsetting feature narrows the list of items to display only those that meet the criteria.

Finally, just as reviewers can create their own patient groups based on patient groups provided by the sponsor, they can also create their own programs, graphs, output, and notes. Programs can either be:

- Copied from the Library provided by the sponsor
- Generated using SAS/PH-Clinical software
- Generated using other SAS software and then included into the Library
- Included from existing SAS libraries
- Included from external files.

Graphs, output, and notes can either be:

- Generated using existing programs with new patient groups
- Generated using existing programs with existing patient groups
- Using new programs with new patient groups
- Using new programs with existing patient groups.

Thus, the Library component is a powerful tool in and of itself during drug development, and also holds a special place in a CANDA system in that it can bring together all analyses performed using any SAS
software product, and can identify data used to perform the analysis.

The Library component can serve as an online repository of all key analyses, graphs, and reports for the study. This feature is useful in developing programs for a number of studies and studies that depend on the same database, thereby saving time and effort.

SAS/PH-Clinical Enhancements
This section discusses a number of recent enhancements to SAS/PH-Clinical software, as well as some enhancements planned for future releases.

Patient Groups
Saved patient group data are a major recent enhancement to the software. In addition to the features discussed earlier, one additional feature useful in CANDA preparation is batch-file-rebuild of saved patient group data. As the data are edited as a result of final audits for errors, the patient group definitions may not need to change but the saved patient group data may need to be updated.

Rather than interactively updating several patient groups, sponsor PAs may find it more efficient to request SAS/PH-Clinical software to build a batch file that contains the code necessary to rebuild these patient groups. Then, the batch file can be scheduled to be run as needed—whether nightly, once a week, once every two weeks, or whatever is needed.

A second recent enhancement for the software involves improvements in the performance of patient group generation. Reviewers who are likely to want to create and save their own patient groups want to have the fastest performance possible. Benchmarking with sponsor data has shown performance improvements from 50% to 85%, with greater improvements seen for more complex studies.

Tasks Enhancements
Once users choose a patient group, they can perform a number of tasks with SAS/PH-Clinical software. As sponsors and FDA reviewers have requested enhancements, the development group has added features to the software. A complete list of recent enhancements is contained in SAS Technical Report P-249 and SAS Technical Report P-248 SAS/PH-Clinical Software: User's Reference, Changes and Enhancements, Release 6.08, but a brief list is given here.

- Ability to sort a browse listing by up to 10 variables instead of only 3
- Enhancements to browse with drilldown to enable drilldown on the patient and visit identifier columns
- Features to control the number of decimal places shown for descriptive statistics
- Ability to specify up to 10 variables for percent change
- Enhanced features for saving single graphs to a GSF
- Enhancements to subsetting in the Library and Overviews components.

In addition, several enhancements to the Administrator Utilities provide new features and reports for study definition, managing users and security groups, and more.

Features available in the (Near) Future
As additional enhancements are requested, the SAS/PH-Clinical software development group will continue to add features. This section discusses enhancements planned for release soon.

A key aspect of CANDA not addressed currently by SAS/PH-Clinical software involves optical images of CRFs. SAS/IMAGE™ software, currently under development, will access images stored in many formats. When this product is available, the browsing features of SAS/PH-Clinical software will be able to access images. Once reviewers find a patient of interest, they can choose a menu item to access a list of images available for the patient and then choose a CRF image to view.

As use of workstations and networked PCs becomes more widespread, the use of client-server models for data management and analysis is increasing. Currently, the data and SAS/PH-Clinical software must reside on the same host, and all processing is performed locally. Future enhancements will allow more diversified processing—users may be able to build
patient groups on the mainframe and then browse the resulting data using the PC processing, for example.

Other features available soon include an improved browse facility, menu-driven export of patient groups to other formats (such as spreadsheets), a patient group review facility that allows users to browse their choices for a patient group, new reports and tables, and enhancements to the variable creation feature.

**Conclusion**

Now that the FDA has officially stated that CANDAs are desirable, all drug development programs should include a CANDA system as a natural extension of data collection and analysis. This should not be an afterthought, but fully integrated with data management and analysis processes beginning with Phase II. Remember that SAS/PH-Clinical software can access data in a variety of data management systems, such as ORACLE®, in addition to SAS data sets.

Planning for submissions in 1995 and beyond should clearly be happening in 1993 (or should have started in 1991). Thus keeping aware of the changing technological trends, such as document management (do you know what SGML is?), should be another responsibility of the CANDA team. Long term thinking assures that new products like SAS/IMAGE software or enhancements to SAS/INSIGHT software are not last minute surprises. Making use of tools such as SAS/PH-Clinical software should become second nature to people in all disciplines, from data management to medical writing.

**References**


**Recommended Reading**


**Contact Addresses**

Dr. J. Mei Mei Ma
Quintiles
P. O. Box 13979
Research Triangle Park, NC 27709

Sandra D. Schlotzhauer
SAS Institute Inc.
SAS Campus Drive
Cary, NC 27513

SAS, SAS/ETS, SAS/INSIGHT, SAS/PH-Clinical, and SAS/IMAGE are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ORACLE is a registered trademark of Oracle Corporation. © indicates USA registration. Other brand and product names are registered trademarks or trademarks of their respective companies.