Abstract
Using Seroquel® sNDA as example, the paper discusses the work and the issues coming along in the process of preparing CTD documents to the FDA when it comes to clinical data integration, table templates, local publishing, efficacy tables, data definition files, etc. The author has introduced AstraZeneca’s commitment to preparing CTD’s and the analysis & reporting standardization, the company’s own electronic library, as well as the ongoing implementation of CDISC standards.

Introduction
Among the three regions, i.e. the United States, Europe, and Japan, the technical requirement for the registration of pharmaceuticals for human use has been harmonized a great deal through the ICH process. The CTD (common technical document) is part of the process that has the most direct impact on the daily work of clinical SAS® professionals. The CTD consists of five modules. The clinical summary in module 2 and the clinical study report in module 5 are where the clinical SAS programming work will contribute.

Throughout the CTD, the display of information should be unambiguous and transparent, to facilitate the review of the basic data and to help a reviewer become quickly oriented to the application contents. Text and tables should be prepared using margins that allow the document to be printed on both A4 paper (E.U. and Japan) and 8.5 x 11” paper (U.S.). The left-hand margin should be significantly large that information is not obscured through binding. Font sizes for text and tables should be of a style and size that are large enough to be easily legible, even after photocopying. Times New Roman, 12-point font is recommended for narrative text. Acronyms and abbreviations should be defined the first time they are used in each module. References should be cited in accordance with the current edition of the Uniform Requirements for Manuscripts Submitted to Biomedical Journals, International Committee of Medical Journal Editors (ICMJE).

The format of the Clinical Study Report (CSR) in module 5 is defined in ICH E3 document. The following is the organization of the CSR with E3 section reference.

• Synopsis (2)
• Study Report (3 to 15)
• Appendices (16)
  – Protocol and amendments (16.1.1)
  – Sample Case Report Forms (16.1.2)
  – List of IECs IRBs (16.1.3) and consent forms (Inspection of Electromagnetic Compatibility and Institutional Review Boards)
  – List/Description of Investigators/Sites (16.1.4)
  – Signatures of Principal Investigator…(16.1.5)
  – List/Patients receiving Test Drug/batch (16.1.6)
  – Randomization Scheme (16.1.7)
  – Audit certificates (16.1.8) and reports
- Statistical Methods (16.1.9)
- Inter-laboratory standardization method (16.1.10)
- Publications based on the study (16.1.11)
- References (16.1.12)
- Discontinued Patients (16.2.1)
- Protocol Deviations (16.2.2)
- Patients excluded from efficacy studies (16.2.3)
- Demographic Data (16.2.4)
- Compliance/drug concentration data (16.2.5)
- Individual Efficacy Response data (16.2.6)
- Adverse Event listings (16.2.7)
- Listing of individual laboratory measurement (16.2.8)

Diagrammatic Representation of the ICH Common Technical Document
**AstraZeneca’s Commitment**

AstraZeneca, an Anglo-Swedish pharmaceutical giant with a strong presence in U.S., is very supportive with regard to the ICH /CTD. As early as in June 2001, AstraZeneca voiced the commitment to CTD to FDA.

“AstraZeneca strongly supports CTD and seeks to utilize it as soon as is feasible. In deciding upon strategies for implementation, AstraZeneca would ask that the following points be taken into account:

- **Commitment** - CTD the future for us all? AstraZeneca urges that all concerned commit completely to CTD. In particular we urge that CTD be adopted as the sole format for the future, not as just an alternative. Also territorial supplements should be kept to a minimum and only requested where CTD really does not provide information essential for local review.

- **Synchronization** - Mandatory implementation CTD is intended as an exercise in global harmonization, therefore the dates for the mandatory use of CTD must be harmonized. Failure to do so will significantly diminish the global impact of CTD, and greatly complicate the implementation of the format within industry, delaying significantly the realization of the benefits CTD will deliver.

- **Mixed dossiers** - formats, old and new scope to submit “mixed dossiers” for an interim period will considerably assist industry in introducing CTD. By mixed dossiers we mean whole modules in old format or CTD. Much Quality and Non-clinical information is already written in old format. Re-writing in CTD format would be costly and add little value. Supplementary applications should be allowed in mixed format where previously submitted data is in old format.

- **Global Electronic Standards** - Truly harmonized electronic standards for all levels of dossiers are essential. Differing standards across regions could cancel out all the savings in time, effort and cost that CTD would otherwise deliver.”

(http://www.fda.gov/ohrms/dockets/dockets/01n_0167/01N-0167_ts00005/)

**Seroquel® sNDA In The CTD Format**

Seroquel® has been one of blockbuster drugs of AstraZeneca since 1997. It has been used to treat schizophrenia. Recently, following the CTD format, supplemental new drug applications (sNDA) had been filed for new indications of bipolar mania and depression. Application for bipolar mania indication had been approved by FDA in 2004, and for bipolar depression was filed at the end of last year. The following is intended to present the work and discuss the issues in the process of preparing the sNDA from SAS programming point of view.
Data integration

When submitting an NDA to authorities, it usually includes the results of multiple clinical studies. Several years ago, ISS and ISE (integrated summary of safety and efficacy) were the common terms for FDA submission. Now they have been replaced with the term CTD. In either case, it is essential to build a set of integrated clinical trial data. When the studies were conducted in different regions or at different times or with different focus, even for the same compound, data structures and associated formats could be very different. AstraZeneca is an international company with numerous clinical trials ongoing; the variation in conducting the trials is not uncommon. Things to be considered for SAS programmers are such as the name of variables, the length of labels, the coding schemes, etc. to give a few examples. When putting them together, the task of integration could be tedious and cumbersome. CDISC (Clinical Data Interchange Standards Consortium) submission data standards (SDS), if implemented, will help lessen the burden in this aspect. While the CDISC standards have not been fully adopted, the clinical information team for Seroquel sNDA in AstraZeneca has programmed a series of SAS programs to generate integrated data sets.

As a matter of fact, AstraZeneca is committed to CDISC standards as well. Several director-level employees have been heavily involved in the CDISC activities, from attending conferences, disseminating the information, communicating between the rule makers and the rule users, to implementing CDISC standards within the company, and so forth. Hopefully very soon, we will not need to write those tiny programs for individual studies in order to do the CTD data integration.

Templates (or prototype) of the tables and listings

There is no specific CTD requirement on the layout of the tables and listings in the Module 5 Clinical Study Report. Thus gives new drug applicants certain leeway to present data as they see fit. However, the non-standardized process of preparing templates/prototype many times causes unnecessary miscommunication in the whole submission process. Except trying to add items of the interest, people do not pay enough attention to the templates because they are merely intermediate steps. Reviewers would rather read carefully the reports with real data than examine the empty table layout and check on the abstract interrelation of their components. But, as we all know, very often templates are important for the SAS programmers to produce the right results.

To improve this process, AstraZeneca is currently implementing standards of A&RT (Analysis and Reporting Transformation), which include a set of standard templates for tables and listings. The purpose of the standardization is to avoid unnecessary variation in the way clinical data are reported across projects at AstraZeneca, and to maximize the efficiencies to be gained from pre-programming of standard data display. These outputs intend to meet the needs of the Regulatory Reviewers to a) minimize questions and requests for information post-submission and b) avoid presentation of superfluous data to control information overload /ensure good readability.
**Tweak existing programs or write new programs**

When preparing the CTD, a decision for SAS programmers to make is to write brand new programs or tweak existing programs of the individual studies to generate tables on combined data. In general, tweaking is supposedly more efficient with respect to saving time and avoiding errors. But it requires programmers be quick in understanding the existing programs and tweaking up as needed. However, in the operation of big pharmaceuticals, such as AstraZeneca’s, personnel change is constant. It is not always easy to find experienced programmers who are capable of quickly mastering the programs written by others. These programs usually have a very long change history by an array of “permanent” and temporary programmers, with regard to their employment status. Moreover, the programs may involve a bunch of macros (necessary or not necessary, see “To macro or not to macro” by Amy Gillespie, Shuping Zhang, et al in PharmaSUG 2005 proceedings) residing in different directories of the computer system, thus make the work more complex. Therefore, writing brand new programs is not necessarily a bad idea provided the timeline of the project allows it. Not surprisingly, the final outcome more likely is a combination.

In order to deal with these issues and smooth out the work stream, AstraZeneca programming group has formed a committee called Standard Tools for Analysis Reporting Team, to explore and develop standard programs to be used across board in the company. Using an analogy of the lunch menu in the company cafeteria, we are trying to provide the customers with standard dishes, and a few today’s specials.

**Publish to the AZ Global Electronic Library**

AstraZeneca’s Global Electronic Library (GEL) is where to store finalized clinical trial documents. The tables and listings produced by SAS programs as part of CTD are first published in GEL. To conform to the presentation format of GEL, which is compatible with the CTD standard, we need to use the SAS Template Procedure to define font, size, margin, style, and background etc. And thanks to the SAS ODS output, the RTF file can be easily created on UNIX®, which is the computer operating system we are using for the bipolar depression sNDA. Thus makes the task of generating in-text tables easier as well.

As SAS becomes more and more powerful, the customers’ desire to have a perfect table has also grown stronger. They once requested the long text in the table such as investigator’s comments be flowed freely within the column (and divided by syllable!), fortunately the flow option in the Report Procedure came to rescue. Now they are asking for special symbols including super/sub script and Greek letters to show up. There are several ways to accommodate this request (see “Using Special Characters and Formats with SAS” by Quan Ren in PharmaSUG 2001 proceedings and “Producing Special Characters in SAS Output in RTF” by Cindy Song in PharmaSUG 2000 proceedings), but they are not very systematic, and sometimes hard to implement when you use an existing print engine to produce the final reports. It is expected that SAS Institute will eventually
help on this one as well. The BYTE function does not seem good enough to make it happen.

**Efficacy tables and data definition files**

Because the drug is on the market, the safety of the drug is less a concern relative to the efficacy, which is the key for the submission. The programs for these efficacy tables usually involve statistical procedures or methodology, which require the programmer to have more knowledge in statistics. They are not as straightforward as calculating percentage or counting the frequency. Statisticians on the project tend to lead the whole analysis and reporting process, and not necessarily have the hands-on knowledge of a particular efficacy program adopted from individual studies. At the same time, thanks to the prosperity of healthcare industry and the popular culture of computer technology, SAS programmers could come from all walks of life and have different backgrounds. But not every SAS programmer is good at the numerical technique and can handle the statistical programming. For the integrated presentation of the efficacy tables, group effort is always needed to ensure the quality and accuracy.

Another task for the submission is preparing the data definition files (DDF) in accordance with FDA 21 CFR Part 11. In AstraZeneca we like to call it “Item 11”. To put it simply, DDF is another form of CRF (Clinical Research Form) for the regulatory authorities with more detailed explanation of certain data. Depending upon which data sets the DDF are based on, the explanation could get pretty complicated. In order to do this, you need to refer to the clinical trial protocols, CRFs, the statistical analysis plans, and never forget their amendments. Sometimes all of these documents are still not enough. As we know, the raw data delivered from the clinical data management are consistent with the annotated CRF fields. They are considered well documented and provide the basis for the programming work in the process. Once the analysis work starts and raw data sets become analysis data sets through some SAS programming, the newly derived variables could be very far from the origin. A lot of times the logic or algorithms used to derive these variables are not documented in detail anywhere else other than being imbedded in the programs that produce them. When an old study and its result had been put aside for a while, it could be very challenging to find out the exact derivation of these variables and then summarize them in the DDF. These derivation could be the decision from several intensive group meetings, or changes that had been going back and forth for a period of time, or a definition of a category of disease which is being controversial in the medical community, or additional criteria for subjects inclusion/exclusion, and so forth. We have a very good program to read in SAS data sets and create the XLS and RTF files for each variable to make the DDF. Associated formats are also generated automatically. What is low tech is the effort to find out the derivation from the programs and then add to the files generated by the program.

**Summary**

When it comes down to concrete SAS programming work for the supplemental new drug application in the pharmaceutical environment, the high level ICH/CTD format has to be supplemented with more detailed data specification, template layouts, standardized
analyses, and powerful software tools. Putting them together will increase the probability of a successful submission.

**Contact Information**
Your comments and questions are valued and encouraged. Contact the author at Charlie Xu

1800 Concord Pike
Wilmington, DE 19850
302-886-4086
charlie.xu@astrazeneca.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. Seroquel is registered trademark of AstraZeneca. ® indicates USA registration.