2013 Annual Conference of Pharmaceutical Industry SAS® Users Group

ABSTRACT

PharmaSUG

CHICAGO, IL
MAY 12-15, 2013
www.pharmasug.org

Nancy Brucken
Operations Chair
inVentiv Health Clinical
OperationsChair@PharmaSUG.org

Craig Brelage
Academic Chair
Eli Lilly and Company
AcademicChair@PharmaSUG.org

ABSTRACTS

The abstracts program provides the details on each of the presentations given at PharmaSUG 2013 including time, place, authors, title, and abstract. Please note that there may be a few last minute changes to the information given in this section. These changes will be noted in 2013 PharmaSUG mobile app and outside the room where the presentation is scheduled on the day of the presentation. The abstracts given on the following pages are ordered by paper number within section. The sections can be found on the following pages.

APPLICATIONS DEVELOPMENT (AD)

Co-Chairs	Company
Vikash Jain	Octagon Research Solutions
Natalie Reynolds	Eli Lilly and Co.

AD01 : GUI-Based Utility Macro for Creating a Version Controlled Project Directory and Copying in Standard Tools and Templates Files
Hisham Madi, INC Research
Anna Maeser, INC Research
Michael Zichy, INC Research
Tuesday, 2:15 PM - 2:35 PM, Location: Chicago Ballroom B

There are many considerable advantages to standardizing a directory structure for all projects/studies. A well-defined project directory structure enhances the organization of study files such as data, SAS programs, output and study documentation. This paper illustrates an approach to creating a standardized directory using a utility macro, a default list of tools and files to be copied into the newly created study directory and an Excel spreadsheet which defines the directory structure. Additionally, the process is driven by a SAS pop-up window interface that collects user-specified options, a management-controlled sponsor list/root directory naming convention, and call system commands to create the necessary subdirectories and copy template files, programs, specs, etc. into the newly created study directory. Lastly, since each component of the setup process is version-controlled, the macro dynamically selects the most recently approved version of the sponsor, tool and directory/subdirectory lists upon each execution of the setup macro.

AD02 : Solving Samurai Sudoku Puzzles -- A First Attempt
John R Gerlach, CSG, Inc.
Wednesday, 9:00 AM - 9:20 AM, Location: Chicago Ballroom B

Imagine a Sudoku puzzle that consists of a 9x9 matrix having about 30 out of 81 cells populated. Now, imagine five such Sudoku puzzles such that a center puzzle is joined to four others at their respective 3x3 corner sub-matrices. These five Sudoku puzzles define a Samurai Sudoku puzzle. The solution to the puzzle is the same: each puzzle having unique values per row, column, and sub-matrix. However, there is an obvious interdependence among the five puzzles that poses a new challenge. This paper explains an expanded version of a SAS solution that used a dynamic cube to solve a regular Sudoku puzzle, by incorporating five cubes to solve the Samurai puzzle.

AD04 : More Power to SAS - Embedding Other Programming Languages in SAS Through SAS/IML® Studio
Max Cherny, GlaxoSmithKline
Tuesday, 8:00 AM - 8:50 AM, Location: Chicago Ballroom B

Embedding code from different programming languages offers many new opportunities for SAS users to further enhance their SAS programs by taking full advantage of the features not yet available in SAS. SAS/IML Studio is a very powerful free SAS software which allows use of various programming languages together with SAS code. These languages include the IML programming language used for complex and fast data analysis, Java, C and R. This paper explains the purpose and the basic functions of SAS/IML Studio. Easy and reusable examples of running SAS/IML, Java and R code in SAS/IML Studio will be provided. These examples can function either as stand-alone programs or as the starting point for further integration of IML, Java or R programming languages into SAS code. The paper will also provide side-by-side comparison of R and SAS syntax.

AD05 : Running OpenCDISC in SAS
Kevin Lee, Cytel
Tuesday, 11:15 AM - 11:35 AM, Location: Chicago Ballroom B

OpenCDISC provides great compliance checks against CDISC outputs like SDTM, ADaM, SEND and Define.xml. OpenCDISC is an easy point-and-click software package for CDISC compliance checks. SAS programmers usually run OpenCDISC as an independent executable file after creating CDISC datasets, but there is a way to execute OpenCDISC from within a SAS program. This means that SAS programmers will be able to create SDTM or ADaM datasets and check for CDISC compliance by running OpenCDISC in the same SAS program. This paper will show how to run OpenCDISC from within a SAS program. It will also show how to call the OpenCDISC compliance check software within a SAS program and how to assign the parameters such as SDTM, SAS Transport file, the location of the source data, the version of compliance checks, the output type, and the location of OpenCDISC compliance check reports.

AD06 : DataSetBuilder - An Application for Creating Edit Check Test Cases Within SAS Data Sets by Non-Programmers
Richard Addy, Rho
Tuesday, 3:30 PM - 3:50 PM, Location: Chicago Ballroom B

Validating edit checks is a necessary but cumbersome process. Individual test cases must be created, and complex checks require many records across multiple data sets. The person validating the checks might not be a SAS programmer, and creating these records is a non-trivial task. Documentation of the test cases can be sparse (or non-existent), and it may be difficult for a second party to confirm the success of the validation. We approached these problems by combining several tools to better handle sections of the validation process. The primary engine of the application is a SAS program

(DataSetBuilder or DSB) which modifies the contents of SAS data sets based on the specifications of a validator. A macro-enhanced Excel spreadsheet is the intermediary between the validator and DSB; this spreadsheet has numerous tools to quickly and efficiently specify test cases. Using Dynamic Data Exchange (DDE), DSB provides feedback to the validator on the test cases it creates, including any problems it encounters. This approach allows the validator to focus on creating test cases, using a familiar tool that does not require any SAS programming proficiency. The SAS program handles the heavy lifting by creating records within the desired data sets and alerting the validator to any potential problems. A second party can quickly review a validator's test cases, or recreate cases on an as-needed basis.

AD08 : Just Press the Button - Generation of SAS Code to Create Analysis Datasets Directly from an SAP - Can It Be Done?
Endri Endri
Rowland Hale, inVentiv Health Clinical
Monday, 3:30 PM - 3:50 PM, Location: Chicago Ballroom B

CDISC is rapidly becoming adopted as the data standard for clinical trials and submissions of clinical trial data to the FDA. Within CDISC, ADaM datasets are an integral part of clinical study analysis and require significant data derivation to fulfill the needs of TLF provision. Since the move towards standardization is nothing new, are we now able to obtain a full study analysis at the mere press of a button? The answer is of course still "no, not yet". The reason is that much still depends on the study design and the structure of the data, the derivation rules or maybe just a small footnote which needs to be printed on a particular table. The paper goes on to explain how it may be possible to "translate" information written in an SAP (such as imputation rules, population flag, baseline flag, etc.) to generate SAS code which will produce the analysis dataset. Using a text analysis algorithm a SAS macro will attempt to identify the requirements and determine the source data needed for creation of the analysis variables. A further challenge is that many pharmaceutical companies and CROs already have their own standard macros which they use to analyze their clinical data and which they will probably wish to continue to use. The code generation solution presented in this paper should bear this in mind and make provision for the continued use of in-house macro libraries as far as possible.

AD09 : A Simple Approach to the Automated Unit Testing of Clinical SAS Macros
Matthew Nizol, United BioSource Corporation
Tuesday, 9:00 AM - 9:20 AM, Location: Chicago Ballroom B

Federal regulations require software used in the analysis of clinical trials to be validated. In the pharmaceutical industry, SAS programs used to generate data sets, tables, listings, and figures are often validated via double programming. However, unit testing is often a more appropriate solution for validating shared macros which are used across multiple studies. Unit testing is the process of executing the smallest component of a software system on a known set of inputs and comparing the resulting

output to a predefined set of expected results. Unit tests provide confidence that code is implemented correctly; they are also a safety net that protects against unintentional changes to software. But, unit tests can be challenging to write and tedious to run. A unit test framework makes the tester's life easier: it provides a library of assertion macros which both standardize and simplify the writing of tests; it provides a means to run multiple tests at once to make regression testing easier after software is changed; and it automatically reports the results of all test runs so that feedback is immediate. This paper will show how to write, in as little code as possible, a simple, maintainable, and robust unit test framework for the testing of clinical SAS macros. This barebones test framework will be useful in its own right, easily validated due to its simplicity, and may also serve as a starting point for an organization to develop a more complex testing solution to meet their own unique needs.

AD10 : A Generic Concept to Handle SDTM (and Other CDISC) Data Sets
Peter Schaefer, BASS, LLC
Tuesday, 4:00 PM - 4:20 PM, Location: Chicago Ballroom B

CDISC-defined standards like SEND, SDTM, or ADaM are specified to be compatible with the SAS version 5 transport file format (.xpt-files). Of course, this compatibility and the widespread use of .SAS throughout the pharmaceutical industry very often makes SAS the tool of choice when it comes to reading and writing study data in any of the CDISC-defined formats. In addition, other tools include SAS transport file access as an extension to their native data file formats, like the widely used open source package R with the SASxport package. However, working with CDISC datasets require more than reading or writing the specific .xpt-file for each domain; for example, compliance requires that certain relationships between data are maintained or created or that controlled terminology is used throughout the data set. The proposed presentation will discuss the concept of a file format independent software package using SDTM data sets as an example. An implementation of this concept as a Web Service demonstrates how programmers can focus on analysis and workflow issues in their program instead of data management. In addition, the additional level of abstraction supports other file formats than SAS transport files as long as CDISC standard requirements are met. With this kind of approach, programmers can take advantage of the many choices when it comes to selecting the appropriate software technology for a program. They are no longer limited by one requirement for file access when other aspects might call for using a different, more appropriate technology.

AD11 : Let SAS Set Up and Track Your Project
Tom Santopoli, Octagon, now part of Accenture
Wayne Zhong, Octagon, now part of Accenture
Monday, 4:00 PM - 4:50 PM, Location: Chicago Ballroom B

When managing the programming activities in a clinical trial, a project tracking document can be a convenient place to store important information. This information may include an inventory of the data sets, tables, listings, and figures (DTLFs), the programs that produce the DTLFs, and the

validation/completion status of the DTLFs. SAS can use this information to dynamically generate program shells and batch scripts before any of the programming work even begins. In addition, SAS can update the validation/completion status of each DTLF by examining the properties and contents of logs and other files in the project folder. In this paper, we will present several very powerful techniques that let SAS do all of this work for us, thereby ensuring high quality completion of the programming tasks in a clinical trial.

AD13 : A SAS-Based MedDRA Coding System
Charley Wu, Alexion Pharmaceuticals
Dona-Lyn Wales, Alexion Pharmaceuticals
Tuesday, 4:30 PM - 4:50 PM, Location: Chicago Ballroom B

A simple and efficient MedDRA coding system is built with SAS. This system has three major components: 1)MedDRA Dictionary and MedDRA Synonym Dictionary. 2) MedDRA coding transaction file. 3) Auto and manual coding processes. The system has the following functions: SAS MedDRA dictionary data loading, MedDRA Synonym Dictionary self-learning and self-cleaning, auto and manual coding, audit trail, versioning, and query generation. Users can code against any version of the MedDRA dictionary. The system is built on a modular basis, so it is easy to maintain and easy to extend its functionalities.

AD15 : A Macro to Batch Submit a List of Programs with Real Time Feedback
Andrew Hansen, Quintiles
Tuesday, 5:00 PM - 5:20 PM, Location: Chicago Ballroom B

Once clinical projects are in the later stages of development or maintenance, it is not uncommon to have to run a large group of programs due to new data or other updates. In recent years there have been multiple SAS papers authored by people in the pharmaceutical industry that describe methods and tools to automate this process. The purpose of this paper is to describe a macro tool for batch submitting a large number of SAS programs that is unique from previous papers in that it provides real-time status feedback to the user as the programs are being executed.

AD16 : Application Interface for Executing a Batch of SAS® Programs and Checking Logs
Sneha Sarmukadam, inVentiv Health Clinical
Monday, 5:00 PM - 5:20 PM, Location: Chicago Ballroom B

The most convenient way to execute multiple SAS programs of interest is by running them together in a batch. Usually, a programmer would think, 'what better way than to create a "batch file" (.bat) for this purpose?' This is definitely the right solution. But, programmers do not usually stop at a batch run. They might also have to check the log files generated, either manually or by using a SAS macro. Since these

two are routine tasks which are performed manually in our clinical domain, what if we combine both of these operations in one application - an application which, with just a few clicks, can carry out all the above activities? To serve this purpose, I have created a simple application - EzeeApp. In this paper, we shall see what EzeeApp is, learn how to use it and how beneficial it is to a SAS programmer.

AD17 : Automated Validation of Third Party Data Imports
Pranav Soanker, PPD Inc
Tuesday, 2:45 PM - 3:05 PM, Location: Chicago Ballroom B

When studies are outsourced to a Contract Research Organization (CRO) by a pharmaceutical company (client) they can include the requirement to use a third party database to configure and report study data. Once a study starts, in-stream changes to the structure of these databases can impact program functionality and/or create delays in development, either of which can be time- and resource-intensive. Therefore at the start of the study, it is critical to have a consensus between the third party vendor and the CRO on the database structure to ensure imports throughout the study are consistent. After an initial agreement is reached, a test transfer is typically performed to validate the import against the agreed-upon specifications. Previously this validation was a manual process that, though critical, was cumbersome and required differences to be noted and compiled separately. The method described within this paper provides an automated process that can be used at study start-up for initial validation, as well as with each new import, to identify mismatches between what was expected and what was received. This process utilizes a data set that holds the structural specifications and compares against the imported data, which can be in SAS data sets, transport files, ASCII files, or Excel spreadsheets. The process results in reports that identify the differences and can be used to clearly communicate them to the vendor.

AD18 : A Simple Interface for Metadata
Magnus Mengelbier, Limelogic AB
Wednesday, 10:45 AM - 11:05 AM, Location: Chicago Ballroom B

The use of metadata in Life Sciences, and specifically clinical trials, has been around from the very beginning. The most common form is to embed information about the trial or analysis in evolving standards, documentation and different forms of specifications. Very seldom are these readable from SAS and other analytics environments. We consider a simple interface that will allow the same details to be captured and managed consistently and provide a few powerful utilities to use it in programming.

AD19 : Automating Validation of Define.xml using SAS®
Prafulla Girase, Biogen Idec Inc
Robert Agostinelli, Sunovion Pharmaceuticals Inc
Tuesday, 9:30 AM - 9:50 AM, Location: Chicago Ballroom B

There are various methods that can be used to validate the define.xml. A community version of OpenCDISC is a popular and efficient way to validate define.xml. However, since it may not cover everything, it is important to supplement it with other validation checks to ensure the accuracy, validity and completeness of your define.xml. Doing this additional validation manually can be cumbersome and time-consuming. Since the define.xml is a machine-readable file, it is possible to automate many of these additional validation checks. In this paper, we discuss an automated SAS-based approach. One important technique that makes this possible is getting define.xml contents into SAS datasets using a SAS XML Map. Once the define.xml contents are transformed into SAS data sets, one can perform a variety of tasks or checks using the flexibility and power of SAS. We developed several custom validation checks in our macro. The checks are related to issues that are within the define.xml itself, and also issues related to inconsistencies between the define.xml and submission XPT files. The paper discusses a method for getting define.xml contents into SAS datasets, all the validation checks that can be performed, and also presents samples of validation output. While we have not provided the entire SAS code in this paper, important snippets of code necessary for someone to get started are provided.

AD20 : Clinical Reporting by Elements
Magnus Mengelbier, Limelogic AB
Wednesday, 8:00 AM - 8:50 AM, Location: Chicago Ballroom B

The reporting of clinical data within Life Sciences is a common and continuous process. The advent and pressures exerted by requests for short reporting timelines, real time data access and multiple channels of information may impact classical methods of data and information management in a very negative way. We consider common industry processes to generate, manage, verify and document summaries, and how this environment can be optimised. The current form of the programs, output formats and quality assurance requirements require several manual tasks that are time-consuming, intensive and may require a larger accumulation of risk of reporting inconsistencies within summaries. We extend new paradigms in generating summaries to keep abreast of requests for real-time access to information while retaining the control over standards, quality and consistency. We also consider how the proposed strategy can be conveniently applicable to efforts in Business Intelligence, information source integration and the next generation of document authoring tools.

AD21 : SDTM Harmonization in the Absence of CDASH Modularized Approach to Domain Programming
Annie Guo, ICON Clinical Research
Wednesday, 10:15 AM - 10:35 AM, Location: Chicago Ballroom B

SDTM is the recommended format for submissions to the FDA. CDASH defines basic standards for the collection of clinical trial data. The two standards streamline the processes of converting CRF data to SDTM format. However, in the absence of CDASH, de-normalized database structure and a lack of use of controlled terms can cause SDTM programming to be time-consuming and errorprone. This paper presents a modularized approach to developing reusable code and facilitating programming. The outcome is reduced programming time and quality output SDTM data.

AD22-SAS : Methods and Application for Determining the Integrity and Veracity of Medical Device Safety Related Data in Social Media
Mark Wolff, SAS
Michael Wallis, SAS
Tuesday, 1:15 PM - 2:05 PM, Location: Chicago Ballroom B

As more individuals, organizations and institutions rely on the Internet for information to support decision-making, the integrity and veracity of those data have become a critical issue. A key area of interest is the applicability and utility of social media data for device safety monitoring. Such data offer a potentially valuable resource for post-marketing device safety surveillance for the industry and regulators. Adoption of these data as a resource has been hampered by concerns related to the accuracy and reliability of these data and a lack of guidance from regulators. Applying the capabilities of SAS Text Analytics, we propose a method for qualifying the veracity of unstructured data collected from Internet sources. Further, we describe its application in post-marketing medical device safety monitoring and signal detection.

AD23-SAS : SAS Drug Development 4.x The Next Generation Platform for Enterprise Analytics
Matt Gross, SAS
Tuesday, 10:15 AM - 11:05 AM, Location: Chicago Ballroom B

For over a decade, SAS has offered the SAS Drug Development solution to the Life Sciences community to help meet the markets needs for data management, analysis, exploration and sharing during the drug development and evaluation phases. In 2012, SAS released its next generation of the SAS Drug Development solution to leverage advances in technology to better meet the clinical and business analytic needs of the Life Sciences industry. Learn how the underlying foundation enables an environment for secure, collaborative work management as well as delivers a globally accessible enterprise analytics environment that enables compliance with industry regulations. Highlighted during the presentation will be the following components and how they differ from the previous 3.x versions:

Web-based SAS Integrated Development Environment ; Workflow and Process Orchestration; the new Data Explorer ; improved user interface; remote SAS session administration; SAS Job Management and Scheduling; secure, auditable, content repository; global access and collaboration capabilities; event-based subscription and notifications.

AD24-SAS : Submitting SAS Code on the Side
Rick Langston, SAS
Wednesday, 9:30 AM - 9:50 AM, Location: Chicago Ballroom B

This paper explains the new DOSUBL function and how it can submit SAS code to run "on the side" while your DATA step is still running. Also explained is how this differs from invoking CALL EXECUTE or invoking the RUN_COMPILE function of FCMP. Several examples are shown that introduce new ways of writing SAS code.

BEYOND THE BASICS (BB)

Co-Chairs	Company
Cecilia Mauldin	Assent
Niraj Pandya	Element Technologies

BB01 : The Hash of Hashes as a "Russian Doll" Structure: Application to Clinical Adverse Events Data Analysis
Joseph Hinson, Princeton, NJ
Monday, 9:15 AM - 10:05 AM, Location: Chicago Ballroom F

The SAS DATA Step hash objects have the unique ability to be generated as a nested structure, popularly known as a "hash-of-hashes". Such structures are particularly well-suited for the processing and regrouping of hierarchical data, by splitting data into distinct tables within tables. In this paper, this approach has been applied to the processing of clinical adverse events data, which often need to be organized into numbers of events within preferred terms within system organ classes within treatment groups. Such complex reorganization has been accomplished with multilevel nested groups of hash objects, all within a single DATA Step, without any sorting or BY-group processing, and with just one pass through the unsorted raw data. It is yet another demonstration of the powerful flexibility and unique opportunities provided by the hash object technique.

BB02 : The Baker Street Irregulars Investigate: Perl Regular Expressions and CDISC
Peter Eberhardt, Fernwood Consulting Group Inc.
Wei (Wesley) Liu, SAS Research and Development Beijing
Monday, 10:15 AM - 11:05 AM, Location: Chicago Ballroom F

A true detective needs the help of a small army of assistants to track down and apprehend the bad guys. Likewise, a good SAS programmer will use a small army of functions to find and fix bad data. In this paper we will show how the small army of regular expressions in SAS can help you. The paper will first explain how regular expressions work, then show how they can be used with CDISC.

BB03 : Not All Equals are Created Equal: Nonstandard Statement Structures in the DATA Step
Arthur Carpenter, CALOXY
Monday, 4:30 PM - 5:20 PM, Location: Chicago Ballroom F

The expression is a standard building block of logical comparisons and assignment statements. Most of us use them so commonly that we do not give them a second thought. But in fact they definitely do

deserve that second thought. A more complete understanding of their construction and execution can greatly expand our ability to more fully take advantage of this fundamental component of the SAS Language. Once we understand the basic form of the expression and how it is used in various statements, we can use this understanding to create statement forms that would otherwise appear to be illegal or just plain wrong. Further and perhaps even more importantly, this deeper understanding can help to prevent us from committing errors in logic.

BB04 : Enhancements to Basic Patient Profiles
Scott Burroughs, GlaxoSmithKline
Tuesday, 10:15 AM - 10:35 AM, Location: Chicago Ballroom F

Patient Data Viewers are becoming more prevalent in the pharmaceutical industry, but not all companies use them nor need them for all situations. Old-fashioned patient profiles still have use in today's industry, but how can they be enhanced? Missing data, bad data, and outliers can affect the output and/or the running of the program. Also, relying on analysis data sets that need to be run first by others can affect timing (vacations, out-of-office, busy, etc.). As always, there are things you can do to make them look prettier in general. This paper will show how to solve these issues and make the program more robust.

BB05 : OpenCDISC: Beyond Point and Click
Frank DiIorio, CodeCrafters, Inc.
Monday, 3:30 PM - 4:20 PM, Location: Chicago Ballroom F

Pinnacle 21's OpenCDISC framework has rapidly gained popularity with programmers who have to validate data compliance with ADaM, SDTM and other CDISC standards. The most common way to use the software is via the user interface. Its clean design and simplicity make it an appealing tool. Given that the software is open source, however, it makes sense to go "under the hood" and explore some of its other capabilities. This presentation discusses some ways the author has ventured beyond the user interface and effectively utilized OpenCDISC. Among these are: running in batch; creating custom configuration files; and workflow required before moving to a new version of the software. Some of the topics require a rudimentary knowledge of XML. Accordingly, we will briefly discuss the structure of the standards' configuration files and some tools that work well with these XML files.

BB06 : Doing the Hashwork: Using the DATA Step Hash Object to Perform Common Clinical Programming Chores
Miles Dunn, Alexion Pharmaceuticals
Monday, 11:15 AM - 11:35 AM, Location: Chicago Ballroom F

The hash table is a widely used data structure in object-oriented programming languages. The purpose of this paper is to give a taste of how the hash object can be used to solve common clinical programming tasks. A hash table, or associative array, is a data structure that maps a key variable or variables to a bucket of associated data values. A SAS format or informat is an example of a simple hash table that maps a single variable to a single associated data value. Although conceptually similar to a format or an array, the hash object is more powerful and flexible. The hash object extends the functionality of SAS formats and informats by allowing both the hash key and the associated data bucket to consist of multiple component parts. With SAS version 9, the hash object and the associated hash iterator object were added to the functionality of base SAS. This paper will show how to use the SAS hash object to merge together subject level data without using sort/merge steps or the SQL procedure. I'll show how to define a hash object and how to load the hash table in a data step. I'll also show how to use object.method syntax to invoke the find, add, replace, first, and output methods against the hash object. I'll present a simple example of how a hash object can be used to store the statistical summary information for a typical clinical lab report.

BB07 : Sharpening Your Skills in Reshaping Data: PROC TRANSPOSE vs. Array Processing
Arthur Li, City of Hope
Tuesday, 8:00 AM - 8:50 AM, Location: Chicago Ballroom F

A common data management task for SAS programmers is transposing data. One of the reasons for performing data transformation is that different statistical procedures require different data shapes. In SAS, two commonly used methods for transposing data are using either the TRANPOSE procedure or array processing in the DATA step. Using PROC TRANSPOSE mainly requires grasping the syntax, and recognizing how to apply different statements and options in PROC TRANSPOSE to different types of data transposition. On the other hand, utilizing array processing in the DATA step requires programmers to understand how the DATA step processes data during the DATA step execution. In this talk, these two methods will be reviewed and compared through various examples.

BB08 : Coding For the Long Haul With Managed Metadata and Process Parameters
Mike Molter, d-Wise Technologies
Monday, 1:15 PM - 2:05 PM, Location: Chicago Ballroom F

How robust is your SAS code? Put another way, as you look through your program, how sensitive is it to changing circumstances? How much code is affected when, for example, the names of data sets to be analyzed or the names of variable within those data sets change? How are those changes expected to be

implemented? In this paper we discuss program optimization and parameter management through the use of metadata. In the wide open, free-text environment of Base SAS, we too often worry more about getting results out the door than producing code that will stand the test of time. We'll learn in this paper how to identify process parameters and discuss programming alternatives that allow us to manage them without having to touch core code. We'll look at SAS metadata tools such as SQL Dictionary tables and PROC CONTENTS, as well as tools for reading and processing metadata, such as CALL EXECUTE. Finally, within our industry, we'll take a brief look at how the Clinical Standards Toolkit puts these methods into practice for CDISC compliance checking. This paper is intended for intermediate-level Base SAS users.

BB09 : Processing MedDRA SMQs: Using Recursive Programming to Handle Hierarchical Data Structures

Paul Stutzman, Axio Research

Tuesday, 9:00 AM - 9:20 AM, Location: Chicago Ballroom F

Programmers sometimes need to process information that is organized hierarchically. Recursive programming is an effective tool for handling these tasks. Recursive programs are programs or functions that call themselves. They are traditionally used in contexts where a basic task can be performed iteratively, as is the case when one processes an operating system's file system by moving through its directories and their subdirectories. The MedDRA adverse event coding system provides a hierarchical way to gather related adverse event Preferred Terms (PTs) into hierarchical groups based on defined medical conditions or areas of interest. These groupings are called Standardized MedDRA Queries (SMQs). Recursive SAS macros are an efficient tool for processing SMQs.

BB10 : Atypical Applications of the UPDATE Statement

John King, Ouachita Clinical Data Services, Inc.

Tuesday, 10:45 AM - 11:35 AM, Location: Chicago Ballroom F

This paper explores uses of the UPDATE statement as it relates to programming clinical trials data. The UPDATE statement is often overlooked as a tool for clinical trials programming; indeed it is often difficult to see how its transaction processing paradigm is directly applicable to programming tables, figures, and listings. Most of us know MERGE and SET, and they along with SQL suffice for most applications. This paper attempts to show through examples how the UPDATE statement can be applied to clinical trials programming problems.

BB11 : Four Useful VBA Utilities for SAS Programmers
David Franklin, TheProgrammersCabin
Monday, 2:15 PM - 2:35 PM, Location: Chicago Ballroom F

While the SAS Programmer can stick strictly to SAS, there are other languages that will assist in making their job a little easier. This paper presents four VBA macro utilities, with code: one that will take all ASCII text files in a directory and save them as Word files; a second that will print out the first two pages of all Word files in a directory; a third that will take all Word files and concatenate them into a single Word file; and finally last but not least, a VBA macro that will search though all the LOG files in a directory and put any issues found into a Word file for review.

BB12 : Bayesian Analysis of Survival Data with the SAS PHREG Procedure
Ryan Brady, Texas A&M
Tuesday, 9:30 AM - 9:50 AM, Location: Chicago Ballroom F

Bayesian analysis has advantages in flexibility and ease of interpretation, but is mathematically complex and computationally intense. Fortunately, the SAS BAYES statement obscures much of the complexity, allowing statisticians and programmers to easily take advantage of that increased flexibility. The BAYES statement is available for four procedures, including PHREG, which is used for analysis of survival data using the Cox proportional hazards model. PROC PHREG uses the partial likelihood as the likelihood, and uses a MCMC Gibbs sampler to generate a posterior distribution. Diagnostic plots are available to determine sampler properties such as mixing, convergence and stationarity.

BB13 : Avoiding SAS Data Set Locks in a Windows Environment
Brandon Graham, PPD
Scott Osowski, PPD
Monday, 2:45 PM - 3:05 PM, Location: Chicago Ballroom F

If you have ever worked in a Windows environment with multiple SAS programmers, and SAS data sets were stored in a common shared folder, you have likely encountered the sometimes frustrating situation of SAS data sets being locked by another user when they are opened for viewing. This paper details a step-by-step method for avoiding SAS data set locks through SAS registry and Windows registry modifications and startup code to alter the default way SAS data sets are opened for viewing. Through this process some basic concepts of the SAS registry and Windows registry settings for SAS files will be discussed.

CODERS CORNER (CC)

Co-Chairs	Company
David Franklin	TheProgrammersCabin.com
Barbara Ross	DataX Ltd.

CC01 : Automating the Footnote Generation
Sonali Garg, Alexion Pharmaceuticals
Catherine DeVerter, Novella Clinical
Monday, 9:15 AM - 9:25 AM, Location: Chicago Ballroom C

Footnotes are always needed in outputs for clinical trial analysis. Due to this need, programs often get littered with too many footnote statements, especially programs which produce multiple outputs. What if a standard footnote could be included in every output produced without having to include a FOOTNOTE statement in each program? What if you could renumber standard footnotes without having to edit your programs? This paper will show how SAS programmers can automate the numbering of footnotes or titles without changing their programs. The method discussed in this paper will make use of the SQL procedure and SASHELP views.

CC02 : Beep, Beep, Beep, Back It Up! A Foolproof Approach to Archiving with No Copying
Kristen Harrington, Rho
Monday, 9:30 AM - 9:40 AM, Location: Chicago Ballroom C

Have you been asked to revise a derivation and then several months later asked to revert back to the previous derivation code? Or when a project ends, were you asked to remove it from the network to free up space? What if archival software is not an option? This paper presents an “archiving” macro (%Archive()) which takes the contents of a directory, including subdirectories, and creates a zip file which contains all files and mirrors the parent’s directory structure. Like magic, your code is saved and easily retrieved! Using an elaborate array of SAS functions, an X command, and WinZip®, the archiving macro determines the execution location and the contents of the subdirectories associated with that location. With virtually no specified parameters, anyone can archive and look like a pro while doing it.

CC03 : Conditional Processing Using the Case Expression in PROC SQL
Kirk Paul Lafler, Software Intelligence Corporation
Monday, 4:00 PM - 4:10 PM, Location: Chicago Ballroom C

The SQL procedure supports conditionally selecting result values from rows in a table (or view) in the form of a case expression. Similar to an IF-THEN construct in the DATA step, a case expression uses one or more WHEN-THEN clause(s) to conditionally process some but not all the rows in a table. An optional ELSE expression can be specified to handle an alternate action should none of the expression(s) identified in the WHEN condition(s) be satisfied. This paper illustrates the basic syntax associated with the two forms of case expression along with examples of each.

CC04 : SAS Logic Coding Made Easy Revisit User-defined Function
Songtao Jiang, Boston Scientific Corporation
Monday, 9:45 AM - 9:55 AM, Location: Chicago Ballroom C

SAS programmers deal with programming logics on a daily basis. Using regular logical operators can solve a complex logic problem, but oftentimes it is not intuitive, error-prone, and hard to read and maintain. Implementation also heavily depends on how well the individual understands the logics and how good the individual programmer's logical thinking is. In this paper, by using SAS user-defined functions, we lay out a universal way of implementing SAS logic without the dependence on an individual;s ability and experience. The solution is straight-forward, error-prove, and clear. Furthermore, these SAS user-defined functions for logical operators can be built into the standard library and utilized by all programmers within an organization. They should be extremely useful to all levels of SAS programmers. To demonstrate the idea, the implementations of these functions are introduced. Two sets of the sample code of moderately complex logic are compared between regular logic implementation and user-defined functions.

CC05 : Can You Export All the Data Sets into Excel for Me? A Small Dynamic Nacro that Does It as Fast as They Think It Should.
Steve Black, WL Gore
Monday, 1:30 PM - 1:40 PM, Location: Chicago Ballroom C

The purpose of this paper is to demonstrate the power of utilizing PROC SQL, the macro language, and the EXCELXP tagset to complete a seemingly simple task of exporting a number of data sets into a single Excel workbook. The secondary purpose of the paper is to briefly explain each section of code moving from basic syntax to a dynamically-driven efficient code which lets SAS do all the hard work for you. The final result is code and understanding which is transferable to any project with a limited number of key strokes.

CC06 : Bring Excel file with SDTM Data in Multiple Sheets to SAS
Mindy Wang, Independent Consultant
Monday, 1:45 PM - 1:55 PM, Location: Chicago Ballroom C

In today's work place, Excel files seem to be the most common files that we deal with. Sometimes we encounter data coming in as an Excel file with multiple spreadsheets. This paper illustrates three approaches to bringing the multiple spreadsheets with SDTM data into SAS, where each sheet becomes a SAS dataset. This first method uses DDE (Dynamic Data Exchange). The second method uses a simple macro program to import the multiple spreadsheets one by one. This third method sets up the Excel file as a SAS library and brings in each sheet as a member in the library. The last method is definitely very convenient and easy to use when there are many sheets involved.

CC07 : Tracking the Use of Standard Programs in Clinical Trials
Adel Salem, VTB
Monday, 10:00 AM - 10:10 AM, Location: Chicago Ballroom C

In pharmaceutical companies, hundreds of programs are used in each trial to generate the needed outputs (tables, listings and graphs). Some of these programs are standard programs, and some of them are custom programs. As a trial programmer, it can be helpful to know which outputs are generated by which programs, and you have also to be sure that you are using the newest version of a given standard program. Some companies use a programming plan (an Excel sheet with a lot of information about each output in the trial, such as title, program name, program type, output name, ID , programmer, reviewer, etc.). It would be nice if there were a tool that could give the programmer information about how many outputs are actually generated by each program, and whether it is a standard or custom program. And if this tool could alert the programmer in the case a newer version of a given standard program exists, this tool would be perfect. The %usestat utility is a SAS macro with some UNIX shell commands that can give this information. It uses UNIX SAS and UNIX shell commands. The intended audience is programmers with knowledge of macro programming and UNIX shell programming.

CC08 : Methods for Deriving COVAL-COVALn in CO Domain
Chunxia Lin, inVentiv Health Clinical
Monday, 10:15 AM - 10:25 AM, Location: Chicago Ballroom C

The COMMENTS (CO) domain contains free-text comments related to data in one or more SDTM domains. COVAL is the variable in CO containing the text of comments. When the text is longer than 200 characters, additional columns COVAL1-COVALn will be generated. This paper shares a couple of methods for splitting the content of comments into snippets of 200 characters long without truncating intact words.

CC09 : Array, Hurray, Array; Consolidate or Expand Your Input Data Stream Using Arrays
William E Benjamin Jr, Owl Computer Consultancy LLC
Monday, 1:15 PM - 1:25 PM, Location: Chicago Ballroom C

You have an input file with one record per month, but need a file with one record per year. But you cannot use the TRANSPOSE procedure because other fields need to be retained or the input file is sparsely populated. The techniques shown here enable you to either consolidate or expand your output data using arrays. Sorted files of data records can be processed as a unit using "BY Variable" groups and building an array of records to process. This technique allows access to all of the data records for a "BY Variable" group and gives the programmer access to the first, last, and all records in between at the same time. This will allow the selection of any data value for the final output record.

CC11 : Combining SAS LIBNAMEs and VBA Macros to Import an Excel File in an Intriguing, Efficient Way
Ajay Gupta, PPD INC
Monday, 2:00 PM - 2:10 PM, Location: Chicago Ballroom C

There are different methods such PROC IMPORT and use of LIBNAME to import Excel files using SAS. While importing the data in SAS, due to technical limitations, these traditional methods might corrupt your data by applying their own automatic formatting. A few examples are: importing a column with mixed data types in a numeric format, or importing regular data containing a hyphen/slash/colon or scientific notation into date/time/scientific formats. The best way to avoid automatic formatting is to import Excel data using a pre-defined format, for example, character. This paper will introduce a unique method where an Excel file is imported using a combination of a Visual Basic for Applications (VBA) macro via DDE in SAS and the SAS LIBNAME statement. This convenient and reliable solution will help SAS programmers/statisticians have better control over the quality of data, and save significant time with minimal coding.

CC13 : Creating a Clinical Summary Table Within a Single DATA Step with the Dynamic Trio: DOSUBL(), Hash Objects, and ODS Objects
Joseph Hinson, Princeton, NJ
Monday, 2:45 PM - 2:55 PM, Location: Chicago Ballroom C

SAS 9.3 provides new features like the DOSUBL function, which allows PROCs to be called within a DATA step, and makes the procedure output immediately available to the same DATA step upon procedure exit. This property of DOSUBL makes it somewhat superior to CALL EXECUTE. SAS 9.3 also features an updated ODS Report Writing Interface (RWI), an object-oriented pre-production tool that provides great flexibility in controlling how data appear in tables and figures. The Hash Object Interface, first introduced with SAS 9, is offered again with a few new methods. These three tools can interact within the DATA step: clinical data can be processed by calling statistical PROCs with DOSUBL, the results organized with hash objects, and the hash data transformed into a final report with the ODS RWI object. This novel approach very much streamlines the intricate task of managing different procedure outputs

to generate a clinical summary report. The present paper is about the production of a demographic summary table, as an example.

CC14 : Graph Your SAS off
Karena Kong, InterMune
Tuesday, 8:30 AM - 8:40 AM, Location: Chicago Ballroom C

This paper demonstrates three different SAS procedures for creating graphs. For illustration purposes, the bubble plot shows the ratio of broadband users (DSL, Cable, Other) ranked by population ("List of countries").The data values of 'Total Subscribers in Millions' and 'Percent Population Online' are annotated on the graph. The three procedures come from SAS/GRAPH - GPLOT, Statistical Graphics (SG) - SGPLOT and the Graphics Template Language (GTL) PROC TEMPLATE with SGRENDER. This paper will discuss the advantages and disadvantages of each one. Based on the comparisons, it recommends which procedure should be used to create a similar graph.

CC15 : Get in Line! How to Make Your Data Alignment Easy!
Ying (Evelyn) Guo, PAREXEL INC
Monday, 10:30 AM - 10:40 AM, Location: Chicago Ballroom C

In data-driven output, we strive for alignment of data that allows for ease of analysis and greater readability. While some outputs need left- or right-alignment, others need decimal-alignment or even alignment with brackets. To make the outputs visually pleasing, column centering is usually preferred, requiring that final datasets be lined up as desired. In order to achieve this, spaces are added to each column. Especially when a column has values of various types, as within a table having means, standard deviations, and ranges in one column, this is time-consuming. Instead of manually adding spaces for each value in that column through trial-and-error, my SAS macro program automatically lines up the values in each column according to a user-specified character. It can automatically search values in a given column, identify the location of the specified character, and add the necessary spaces before or after the value to meet the alignment rules. Use of this macro will significantly shorten programming time usually spent on aesthetics of outputs, thus helping to increase TFL production efficiency.

CC16 : An Alternate Log Axis using SAS PROC GPLOT
Lucius Reinbolt, DataCeutics, Inc.
Tuesday, 8:45 AM - 8:55 AM, Location: Chicago Ballroom C

It is typical in the pharmaceutical industry to plot pharmacokinetic data using a log axis. Pharmacokinetic data is assumed to be log normally-distributed. Setting the AXIS statement to use a log base of 10 allows SAS PROC GPLOT to create semi-log plots. These plots use intervals that correspond to the orders of magnitude as opposed to the standard linear scale. Each unit increase on the log scale represents an

exponential increase in quantity for the given base. This gives the graph an appearance of having uneven tick marks. Also, these log plots will often force the upper or lower axis value to go up to the next power of 10, potentially leaving a lot of dead space in the graph. This can cause the range of values to look condensed and make it difficult to see a pattern. Sometimes it is necessary to remove this dead space while maintaining the log appearance of the graph. A solution to this is to take the log of the data and plot it as you would regular continuous data, then customize the axis to show the tick marks as a log scale would present them. This gives the programmer more control over the range of values plotted while maintaining the characteristics of a log scale.

CC17 : Numeric and Decimal Place Alignment in RTF Files with Non-Monospaced Fonts
Gary Moore, Moore Computing Services, Inc.
Monday, 11:00 AM - 11:10 AM, Location: Chicago Ballroom C

Summary tables in RTF format for population data, laboratory data, vitals & they all have the same problem. Everyone wants that RTF table to look as sharp as possible. But, getting the alignment of numeric data using proportional fonts can drive you crazy. This paper presents simple approaches to solving this problem using column styles and RTF codes.

CC18 : Creating a Batch Command File for Executing SAS with Dynamic and Custom System Options
Gary Moore, Moore Computing Services, Inc.
Tuesday, 10:15 AM - 10:25 AM, Location: Chicago Ballroom C

You would like to customize your SAS environment, but as an employee or contractor without Admin privileges, you can't make the necessary changes to the configuration files in !SASROOT. Or perhaps you would like to dynamically assign the destination of the.LOG and .LST file. How do I accomplish these with only user privileges? With a little knowledge of the Batch Command Language, you can create a batch file that will invoke SAS with the custom system options you need.

CC19 : PRXCHANGE: Accept No Substitutions
Kenneth Borowiak, PPD, Inc
Monday, 3:30 PM - 3:40 PM, Location: Chicago Ballroom C

SAS provides a variety of functions for removing and replacing text, such as COMPRESS, TRANSLATE and TRANWRD. However, when the replacement is conditional upon the text around the string, the logic can become long and difficult to follow. The PRXCHANGE function is ideal for complicated text replacements, as it leverages the power of regular expressions. The PRXCHANGE function not only encapsulates the functionality of traditional character string functions, but exceeds them because of the tremendous flexibility afforded by concepts such as predefined and user-defined character classes, capture buffers, and positive and negative look-arounds. Only Base SASis required. This paper should be

of interest to programmers without any experience with the PRX functions, as well as those well-versed with regular expressions.

CC20 : When Asked for Subject Incidences, Go Ahead and FREQ OUT
Rod Norman, inVentiv Health Clinical
Monday, 3:00 PM - 3:10 PM, Location: Chicago Ballroom C

A frequent task of the clinical programmer is to provide subject incidences or counts of specific values for relevant study variables. For complex ADaM datasets, table specifications may request counts structured from both the differing values of individual variables as well as values of different variables found with the dataset. To perform such counts, programmers frequently rely on restructuring the data so that subject incidences can be more directly calculated. Layers of SQL SELECT statements or macro looping are commonly employed. In this paper, I show that use of the FREQ procedure with ODS OUTPUT can more directly provide a data set that is relevant to obtaining incidences. Drawing on an example of summarizing adverse events from an ADaM-like data set, I show how one run of PROC FREQ with a MULTIPLE specification in the TABLE statement can generate a data set that is structured for a direct counting of subject incidences of adverse events organized by Medical Dictionary for Regulatory Affairs (MedDRA)terms incorporating customized MedDRA queries (CQs), Common Terminology Criteria (CTCAE) and a seriousness classification.

CC21 : Dynamic Project Set-up and Programming Using SAS Automatic Macro Variables and Environment Variables
Gary Moore, Moore Computing Services, Inc.
Monday, 2:30 PM - 2:40 PM, Location: Chicago Ballroom C

Often, we find ourselves working on multiple projects with multiple programmers. The projects are located in various directories. Some programmers prefer a particular editor, while others prefer working in the Display Manager System. All of the work, no matter how it was developed, will need to come together at completion and be updated. In this working environment, it is important to have a flexible way to assign library names and create programs that will work no matter where the project is located or which programmer is working on it. This paper presents dynamic code using SAS automatic macro variables and environment variables that help resolve the conflicts that can occur in this working environment.

CC22 : A Closer Look at PROC SQL's FEEDBACK Option
Kenneth Borowiak, PPD, Inc.
Monday, 3:45 PM - 3:55 PM, Location: Chicago Ballroom C

The FEEDBACK option on the PROC SQL statement controls whether an expanded or transformed version of a query using terse notations is written to the SAS log. This paper will review some of the

documented features of this option and provide additional programming conventions that are explicitly stated when the option is enabled. It will be shown that the FEEDBACK option is an invaluable tool for understanding how PROC SQL processes a query and how it can be used as a code generator.

CC23 : Extending the PRX Functions with PROC FCMP
Kunal Agnihotri, PPD, Inc.
Kenneth Borowiak, PPD, Inc.
Monday, 10:45 AM - 10:55 AM, Location: Chicago Ballroom C

Beginning with SAS Version 9, users have been afforded the ability to perform complex string matching, replacement and extraction with Perl regular expressions via the PRX functions and call routines. We present two user-written functions with PROC FCMP which extend the functionality of PRX. The first function allows for two additional arguments in PRXMATCH, one which allows the users to control which position in the string to begin the match and the other to control whether the beginning or ending location of the string is returned. The second function expands PRXPOSN, which allows the user to extract a capture buffer from a source in single call, as opposed to a sequence of calls to PRXPARSE, PRXMATCH and PRXPOSN.

CC24 : Combining the First Page of Multiple RTF Outputs in SAS using Bookmarks and VBA Macros
Ajay Gupta, PPD Inc
Monday, 2:15 PM - 2:25 PM, Location: Chicago Ballroom C

In order to expedite the review process and save programming time, it would be helpful if the programmer could create a document containing the first page of all the tables and listings. This will help both programmers and reviewers to review the format of all tables and listings at once, and if required, make necessary updates. Traditionally, the programmer/reviewer has to open each table and listing to check the report formatting, which can be tedious and prone to error, especially if there are too many reports. Unfortunately, SAS does not provide a function to get the first page of an RTF report. In my previous paper, I introduced a method for combining multiple RTF outputs. This paper will introduce a method for getting the first page of each RTF document using a bookmark, and later, for combining the multiple documents (all first pages from respective reports) into one document in SAS using Word Basic commands and VBA macros via DDE (Dynamic Data Exchange). If needed, these functions can be used independently.

CC25 : Add a Little Magic to Your Joins
Kirk Paul Lafler, Software Intelligence Corporation
Monday, 4:15 PM - 4:25 PM, Location: Chicago Ballroom C

To achieve the best possible performance when joining two or more tables in the SQL procedure, a few considerations should be kept in mind. This presentation explores options that can be used to influence the type of join algorithm selected (i.e., step-loop, sort-merge, index, and hash) by the optimizer. Attendees learn how to add a little magic with MAGIC=101, MAGIC=102, MAGIC=103, IDXWHERE=Yes, and BUFFERSIZE= options to influence the SQL optimizer to achieve the best possible performance when joining tables.

CC26 : Automating the Labeling of the X- Axis
Sanjiv Ramalingam, Vertex Pharmaceuticals Inc.
Tuesday, 9:00 AM - 9:10 AM, Location: Chicago Ballroom C

Labeling of the X-axis usually involves a tedious AXIS statement specifying tick marks and typing labels for each of the time points. A methodology for implementing and automating the generation of the AXIS statement for X-axis labeling with real time representation of the X-axis is discussed.

CC27 : Creating Customized Spaghetti Plots
Sanjiv Ramalingam, Vertex Pharmaceuticals Inc.
Tuesday, 9:15 AM - 9:25 AM, Location: Chicago Ballroom C

Line plots of individual subjects grouped by treatment for a parameter of interest can be used to observe the parameter of interest over time. The individual subject line plots are distinguished by unique symbols for each subject, and the treatment associated with each subject is distinguished by different line styles. The plot also includes a double legend for subject identification and for associated treatments. A graph with such features is not directly implementable by SAS graphical procedures. A methodology including the use of the annotate feature for implementing such a graph is discussed in the paper.

CC28 : How MEAN is T-test?
Naina Pandurangi, inVentiv Health Clinical
Tuesday, 8:00 AM - 8:10 AM, Location: Chicago Ballroom C

Statistical analysis is a blend of pure statistics and pure programming. Often the written description of the required analysis that is framed by the statistician may sound tricky to program and may make us fetch for the SAS/STAT procedures. But sometimes, the underlying requirement can be easily derived using the regular basic non-STAT, i.e., non-statistical SAS procedures! One such case is where the t-test

analysis is to be performed on one sample with the null hypothesis of mean=0. This paper is going to explain some steps on the bridge between the STAT and non-STAT SAS procedures with the help of two situations where the requirement demands one sample t-test analysis and which can be met using PROC MEANS.

CC29 : Handling Dynamic Variable Types in SAS®
Venkat Lajapathirajan, PPD Inc.
Tuesday, 8:15 AM - 8:25 AM, Location: Chicago Ballroom C

The variable type plays a vital role in SAS datasets; it can be either numeric or character. If the type changes on each data transfer, programmers will need to revise analysis dataset specifications and programs multiple times, reducing efficiency, burning additional hours and affecting gross profit. This paper shows how to handle dynamic variable types through the macro %ConvertType. This macro converts the specified input variable(s) to character type, and the revised variable(s) can be referenced in further developments.

CC30 : Useful Tips for Handling and Creating Special Characters in SAS
Bob Hull, SynteractHCR, Inc.
Robert Howard, Veridical Solutions
Tuesday, 9:30 AM - 9:40 AM, Location: Chicago Ballroom C

This paper will discuss various ways of creating and dealing with special characters in SAS. Many people experience difficulty when reading in Excel files and discovering that strange "boxes" appear in the data. What these are and how they can be dealt with will be discussed. Can special characters be saved in the SAS program? How can these characters be typed if they aren't on the keyboard? We will also provide examples on how to include special characters like Greek letters, less than or equal to symbols, and registered trademark into your SAS programs and RTF output. This paper will help you better understand some ways that special characters can be used within SAS.

CC31 : A Quick Patient Profile: Combining External Data with EDC-Generated Subject CRF
Titania Roberson, Grifols Therapeutics Inc
Yang Han, Grifols Therapeutics Inc
Tuesday, 9:45 AM - 9:55 AM, Location: Chicago Ballroom C

Patient profiles are useful for the review of individual subject data. Most EDC systems can now generate subject eCRFs with the click of a button. These subject eCRFs could be used as the patient profile. The only missing component to a complete patient profile is the external data. This paper explains how to combine 2 PDFs (the subject eCRF generated from the EDC system and a PDF of external vendor data) which lessens the programming effort usually required for producing a full patient profile.

CC32 : Using SAS to Locate and Rename External Files
Lu Gan, PPD, Inc
Monday, 11:15 AM - 11:25 AM, Location: Chicago Ballroom C

When reading external data into SAS, a program needs to specify the data file location and file name in the related import statements. If the raw data file is named with a timestamp, the import program will require an extra step of updating the file name every time before it's run. This will be a hassle if the import is needed on a routine schedule and/or there are multiple external data files. Therefore, a SAS program can be written to eliminate the repeating manual update. This approach includes two steps, first searching for the latest data file in the specific file directory, then renaming it by removing the specific timestamp. After execution of the above steps, the import program will be able to read in the data files directly. The examples and syntax illustrated use SAS 9.2. This paper assumes that the reader has a basic understanding of DATA step programming and the macro language.

CC33 : Macro for Conducting Consistency Checks
Walter Hufford, Novartis Pharmaceuticals Corporation
Tuesday, 10:30 AM - 10:40 AM, Location: Chicago Ballroom C

Pharmaceutical organizations are transitioning from legacy data standards to the CDISC Study Data Tabulations Model (SDTM) and the Analysis Data Model (ADaM) in anticipation of the FDA mandating their use in electronic submissions. Moving to these new data standards is fraught with challenges and there is usually a steep learning curve followed by a bumpy implementation. Although most Data Managers and Statistical Programmers working in the pharmaceutical industry have been exposed to these data models, most are not experts in them yet. This often leads to inconsistent data set and variable attributes - the nemesis of all Statistical Programmers. PROC COMPARE is a useful tool when comparing 2 datasets for inconsistencies; however, it does not provide the ability to compare more than 2 datasets simultaneously. This paper presents a generic macro designed to identify data set and variable attribute discrepancies across N datasets simultaneously and export those discrepancies into a user friendly CSV format for quick review and resolution.

CC34 : Using SAS Driver Programs to Automate Workflows and Respond to the Unexpected
Brit Minor, REGISTRAT-MAPI
Tuesday, 10:45 AM - 10:55 AM, Location: Chicago Ballroom C

A driver is a program that runs another program. Typical uses for SAS drivers are to 1. run several programs that must be run in a certain sequence. 2. parse program logs looking for potential problems. Drivers can perform other routine tasks, though, such as: sending success/failure emails, logging each run of a program to an activity log, and uploading program results to a study portal. Thus, driver programs can automate workflows, improve response to failures, and enforce best practices. This paper will describe SAS macros, techniques, and batch command files for writing driver programs.

CC35 : A One Line Method for Extracting a Substring from a String using PRX
Joel Campbell, Advanced Analytics, CRO
Tuesday, 11:00 AM - 11:10 AM, Location: Chicago Ballroom C

Perl Regular Expressions (PRX) are powerful tools available in SAS® and many other programming languages and utilities which allow precise and flexible pattern matching. The SAS 9 Language Reference provides examples of extracting a substring from a string using PRX functions that require, at minimum, 5 elements, and are potentially difficult to follow for someone looking for a basic example. In this paper, I'll demonstrate a single line method for extracting a substring from a string using only the PRXCHANGE function.

CC36 : Don't Get Blindsided by PROC COMPARE
Roger Muller, Data-to-Events.com
Joshua Horstman, Nested Loop Consulting
Tuesday, 11:15 AM - 11:25 AM, Location: Chicago Ballroom C

"NOTE: No unequal values were found. All values compared are exactly equal." In the clinical trial world, that message is the holy grail for the programmer tasked with independently replicating a production data set to ensure its correctness. Such a validation effort typically culminates in a call to PROC COMPARE to ascertain whether the production dataset matches the replicated one. It is often assumed that this message means the job is done. Unfortunately, it is not so simple. The unwary programmer may later discover that significant discrepancies slipped through. This paper will briefly overview some common pitfalls in the use of PROC COMPARE and explain how to avoid them.

CC37-SAS : Turn Your Plain Report into a Painted Report Using ODS Styles
Cynthia Zender, SAS
Monday, 4:30 PM - 5:20 PM, Location: Chicago Ballroom C

In order to use STYLE= statement level overrides, you have to understand what pieces or areas of PRINT, REPORT and TABULATE output you can change. And then you have to understand how and where in your procedure syntax you use the STYLE= override syntax. Last, but not least, you have to know the names of the style attributes that you want to change. This presentation illustrates how to use STYLE= overrides with PRINT, REPORT and TABULATE with concrete examples. As the examples move from the simple to the complex, you will learn how to change fonts, add text decoration, alter the interior table lines, perform traffic-lighting, and insert images into your ODS output files using some ODS magic to improve your reports.

CC38 : A strategy for Ensuring Non-Estimable Confidence Intervals Having Equal Lower and Upper Confidence Limits are Displayed Correctly

Claudia Jimenez-Castro, inVentiv Health Clinical

Tuesday, 11:30 AM - 11:40 AM, Location: Chicago Ballroom C

This paper examines a particular case in which the scientific notation used by SAS to represent numeric variables resulted in an incorrect representation of non-estimable confidence intervals in an efficacy table. For these intervals, both lower and upper confidence limits were equal. Intervals with both values equal are non-estimable, and were coded to be displayed as such. To determine non-estimable values, the lower limit was subtracted from its corresponding upper limit. However, due to the floating point arithmetic used by SAS, subtracting “a” from “a” yielded a result different from zero. This paper describes the reason behind the error and suggests a strategy to fix the issue.

DATA STANDARDS (DS)

Co-Chairs	Company
Amie Bissonett	inVentiv Health Clinical
Cindy Song	Sanofi

DS02 : Leveraging SDTM Standards to Cut Datasets at Any Visit
Anthony L. Feliu, Genzyme
Stephen W. Lyons, Genzyme
Monday, 1:45 PM - 2:05 PM, Location: Los Angeles/Miami

Clinical trials with complex design or long duration often include interim milestones for evaluation of safety and efficacy. Since clinical databases are seldom configured to deliver other than full extracts, the responsibility for subsetting data invariably falls to the statistical programmer. In this paper, the authors walk through the process of specifying, programming, and validating a generic macro to cut SDTM datasets at any scheduled visit. Three considerations dominated our ultimate design: (a) the tool could be used on any protocol; (b) the tool would first determine the timing variables present a given domain and then apply appropriate cascade logic; and (c) a diagnostic footprint would allow for human review of the decisions made by the tool.

DS03 : Programming Validation Tips for SDTM prior to Using OpenCDISC Validator
Dany Guerendo, DIA, PhUSE
Monday, 2:15 PM - 3:05 PM, Location: Los Angeles/Miami

In the years I have been working with CDISC standards, primarily SDTM but often AdAM as well, validating the domains created has proven to be challenging for programmers new to the standards. This paper provides tips and techniques I developed over the years for validating domains prior to running into OpenCDISC or WebSDM. My thought process was that OpenCDISC is there as a free tool to check for SDTM mapping consistency. We still need to review our data as thoroughly as we can prior to using any validation tool. The idea is to help programmers starting with the standards identify where they should look for possible issues. I provide short and easy to use macros that can easily be recreated or updated to perform daily validations of domains via parallel programming. These ideas can also be used to self-validate ones' code if no official validation programmer is available. Although I worked mostly with version 3.1.2 of the SDTM Implementation Guide, these tips work well with version 3.1.3 which is now available. These programming tips were applied in SAS interactive (Window base d version) version 9.2 and 9.3 as well as SAS Enterprise Guide version 3.1 and 5.3.

DS04 : Implementation Considerations for PARAM/PARAMCD using ADaM BDS
Karl Miller, inVentiv Health Clinical
Joseph Hantsch, inVentiv Health Clinical
Monday, 1:15 PM - 1:35 PM, Location: Los Angeles/Miami

One of the key variables to the ADaM BDS dataset is the PARAM variable which describes the analysis parameter and is directly usable in Clinical Study Report displays or analysis table headings. Current compliance consists of stuffing all essential information including the full test name, specimen type, unit of measurement, etc. into PARAM and generating a system to squeeze this information into the eight characters of PARAMCD. Such PARAM variables frequently become too unwieldy for column headings in the Clinical Study Report and analysis tables, and require an arcane process to generate a uniquely matched PARAMCD. Consequently, when too much data is stuffed into only eight characters, an incomprehensible value is the result. Also, dataset inquiries are hampered by an overly complex PARAM variable value. Along with impacting the PARAMCD variable, such overstuffed PARAM variables carry potential issues into the programming of the analysis itself. We examine the current implementation as well as a functional alternative, non-standard ADaM implementation for generation of the PARAM/PARAMCD variables. Our alternative method reduces PARAM to just the full test name and the specified units, uses the PARCATy variables for any qualifying information, and keeps PARAMCD as a simple abbreviation of the test name reserving the eighth or last character for encoding which of the alternative units is reported. Although our method does not conform to current ADaM Guidance, it maximizes the functionality and transparency of the programming and analysis.

DS05 : Building Traceability for End Points in Analysis Datasets Using the SRCDOM, SRCVAR, and SRCSEQ Triplet
Xiangchen Cui, Vertex Pharmaceuticals Incorporated
Tathabbai Pakalapati, Cytel Inc.
Qunming Dong, Vertex Pharmaceuticals Incorporated
Monday, 3:30 PM - 4:20 PM, Location: Los Angeles/Miami

To be compliant with the ADaM Implementation Guide V1.0, traceability features should be incorporated to the maximum possible extent in study analysis datasets. The SRCDOM, SRCVAR, and SRCSEQ triplet is used to establish data point traceability in ADaM datasets. It can facilitate transparency in FDA submission data, build confidence in analysis results, help efficient programming validation, speed up the overall review progress by FDA reviewers, and build a good relationship with FDA reviewers. This paper provides various examples of applying the SRCDOM, SRCVAR, and SRCSEQ triplet to establish traceability in efficacy ADaM datasets from the cystic fibrosis therapeutic area, and shows the art of applying the triplet to different scenarios.

DS06 : Designing and Tuning ADaM Datasets
Songhui Zhu, K&L Consulting Services
Tuesday, 3:30 PM - 3:50 PM, Location: Los Angeles/Miami

The developers/authors of CDISC ADaM Model and ADaM IG made an enormous effort to give detailed guidance on implementing ADaM for clinical study data. However due to the complexity of clinical trials, they also give the users some flexibility while implementing ADaM standards. Even so, in practice, creating CDISC-compliant ADaM datasets is not easy. In some cases, a bad choice in the early stages may result in datasets that are not CDISC-compliant and almost impossible to make CDISC-compliant in the end. This paper will present author's practice on some critical choices while designing the data structures of ADaM datasets. The topics include: 1) whether to populate CRITy variables in each row or only in the qualified rows, 2) whether to split data sets or categorize parameters in one data set, 3) whether to utilize CRITy or add rows, 4) the relation between ANLzzFL and ABLFL, 5) mapping between AVAL and AVALC, 6) whether to dump variable in ADSL or create more data sets, and 7) whether to derive everything in an ADaM dataset based on SDTM data only or based on SDTM data and other ADaM datasets.

DS07 : Macro %D_ADSL: Automating ADSL Creation from Metadata Files
Jian Hua (Daniel) Huang, Celgene
Monday, 10:15 AM - 10:35 AM, Location: Los Angeles/Miami

As CDISC standards are being more broadly accepted in clinical trials, many tools have been developed for automating SDTM creation. Conversely, there are not many tools available for creating ADaM datasets due to two reasons: First, the ADaM data structures are more flexible than SDTM; and second, the ADaM derivations are more complex and study-specific. Both reasons make it difficult to handle ADaM derivations within one macro program. Fortunately, we can define the ADaM structure and algorithm in the metadata, and then use the metadata to create ADaM data sets. A macro program (called %D_ADSL) has been developed to automate ADSL creation by reading the information from its metadata file. The macro first creates each variable in a pre-specified order, from a simple copy of the SDTM variable to the variables with most complicated derivation. Then it reads all variable attributes (i.e. name, label, length, and controlled terms) from the metadata into ADSL. In addition to creating the ADSL dataset, the macro can also output SAS code for self-review and for further modification. Because this macro uses the metadata to create ADSL, it overcomes the challenge of automatically creating ADSL while handling the complexity of derivation. More importantly, the macro improves the traceability of ADSL derivations and ensures consistency between ADSL and the metadata.

DS08 : Data Standards Will Be Required: Challenges for Medical Device Submissions
Carey Smoak, Roche Molecular Systems, Inc.
Kit Howard, Kestrel Consultants
Fred Wood, Octagon Research Solution, Inc.
Rhonda Facile, CDISC
Tuesday, 10:45 AM - 11:35 AM, Location: Los Angeles/Miami

On October 25, 2012, the FDA made a clear statement at the annual CDISC Interchange in Baltimore, Maryland that electronic standards such as those developed by CDISC will be required. This has particular implications for medical device submissions due to the lack of familiarity with CDISC standards by sponsors and the FDA. While the FDA does accept SDTM domains for medical device submissions, there is currently no consistent data standards and/or requirements for medical device submissions. CDRH is required to develop standards for submissions, and the work of the Device Team will be important in developing these standards. The Device Team (in cooperation with the FDA) has developed seven new SDTM domains. However, much work remains to define the standards for submission of medical device data in an electronic format. To that end, a CDISC pilot with CDRH is planned. This paper will focus on the challenges for sponsors and the FDA in defining these standards.

DS09 : From Standards that Cost To Standards that Save: Cost-Effective Standards Implementation
Jeffrey Abolafia, Rho Inc.
Frank DiIorio, CodeCrafters Inc.
Monday, 4:30 PM - 5:20 PM, Location: Los Angeles/Miami

Recent FDA guidances have made CDISC models the de facto standard for submissions. These standards are the foundation for systems that can improve workflow throughout the project life cycle. They create uniformity of metadata and data structures, which, in turn, provides an opportunity to write generalized, reusable code. When implemented properly, standards have tremendous potential for significant cost savings throughout the project life cycle. However, when implemented poorly, producing CDISC deliverables can actually increase the time and cost of drug development. This paper outlines a new paradigm for cost-effective implementation of CDISC standards that is based on three principles: 1) Adopt a “Tables-first” approach: starting with the end instead of the beginning; 2)Extend the use of standards upstream to the protocol and downstream, to analysis and reporting; and 3) Create a data standards plan as part of the IND or IDE that that is part of the product development strategy and that covers the entire life cycle of product development.

DS11 : Standardizing the Standards: A Road Map for Establishing, Implementing and Unifying Standards in Your Organization
Amy Caison, PPD
Jhelum Naik, PPD
Tammy Jackson, PPD
Tuesday, 1:15 PM - 2:05 PM, Location: Los Angeles/Miami

Electronic data submission is the future of clinical trials and reviewers will soon have the statutory authority to reject submissions that do not conform to expectations (e.g., via PDUFA V, section XII in the US). Data standards comprise the core of these expectations, with SDTM standards as the centerpiece. The extensive body of existing guidance is undergoing rapid development and expansion, including the addition of therapeutic area standards which supplement and extend the existing SDTM implementation guidance. As the deadline for standardized submissions approaches, it is incumbent upon all organizations involved in the pharmaceutical industry to internalize the standardization of data collection, transformation, and analysis to ensure compliance with emerging regulatory requirements. This paper explores the challenges of developing, implementing, and maintaining organizational SDTM implementation standards that are flexible, current and harmonized with industry standards. We present an example case illustrating how a standards initiative can be established and managed. The case highlights the fundamental need for organizational commitment, the value added by SDTM and functional area subject matter experts, and the importance of an extensible infrastructure that provides process and content support to study teams. This infrastructure encompasses standard CRFs, comprehensive mapping specifications, programming tools, and an extensive professionalization of standards knowledge including in-depth training and standards implementation support for study teams. The case will discuss this organizational standards initiative, outlining the challenges and opportunities along with providing a roadmap for ongoing development to maintain currency, while extending and embedding the role of standards in the organization.

DS12 : The Y2K17 Bug! Using a System of Metadata to Respond to PDUFA V Requirements
Vincent Amoruccio, Alexion Pharmaceuticals
Tuesday, 2:15 PM - 2:35 PM, Location: Los Angeles/Miami

On July 9, 2012, President Obama signed into law the Food and Drug Administration Safety and Innovation Act (FDASIA). As part of FDASIA, the Prescription Drug User Fee Act (PDUFA) was reauthorized for the fifth time. PDUFA V requires submitting standardized data to the FDA. If the data submitted is not in a standardized format, then PDUFA V gives the FDA the authority to refuse the submission. At the 2012 CDISC Interchange the FDAs CDER and CBER Divisions reiterated the sentiments of PDUFA V and the eStudy Data Guidance with a very clear message: Standards will be mandated by 2017! While the standards are still in draft form, we have been enlightened by the FDA and CDISC of what to expect. First, the FDA has already announced its expectation to only receive electronic submissions. Second, the FDA has strongly alluded to mandating SDTM and ADaM standards. Third, we anticipate the requirement of metadata as a standard component of SDTM and ADaM. At Alexion

Pharmaceuticals, we have been using a platform-neutral metadata system as a standard way to produce clinical data output for submissions to regulatory agencies such as the FDA. This home-grown cost-effective solution uses basic SAS software and will enable us to easily respond to the PDUFA V requirements. In this paper, I will introduce this simple system and explain how it can be immediately used for regulatory submissions to the FDA and will meet the expected PDUFA V requirements.

DS13 : Experiences in Preparing Summary-Level Clinical Site Data within NDA Submissions for FDA Inspection Planning
Xiangchen Cui, Vertex Pharmaceuticals Incorporated
Tuesday, 4:00 PM - 4:50 PM, Location: Los Angeles/Miami

The Center for Drug Evaluation and Research (CDER) issued a draft guidance in December 2012, which urges sponsors to submit a clinical dataset that describes and summarizes the characteristics and outcomes of a clinical investigation at the level of the individual study site (summary-level clinical site data). The agency uses it to facilitate use of a risk-based approach for the timely identification of clinical investigator sites for on-site inspection by CDER during the review of marketing applications. CDER approved two NDAs from Vertex Pharmaceuticals Incorporated (hepatitis C and cystic fibrosis) in 2011 and 2012, respectively. This paper provides two different examples of preparing summary-level clinical site data for these two drug NDAs. The sharing of hands-on experiences in this paper is intended to assist readers in fully preparing summary-level clinical site data for their NDA submissions.

DS14 : Considerations in the Use of Timing Variables in Submitting SDTM-Compliant Datasets
Jerry Salyers, Octagon Research Solutions-Now a part of Accenture Life Sciences
Richard Lewis, Octagon Research Solutions-Now a part of Accenture Life Sciences
Fred Wood, Octagon Research Solutions-Now a part of Accenture Life Sciences
Wednesday, 10:15 AM - 11:05 AM, Location: Los Angeles/Miami

Often, the appropriate use and population of Timing variables can present many challenges for sponsors when converting their operational database or legacy data to an SDTM-compliant format. This presentation will discuss some of the issues we commonly see in three general areas. The first of these occurs in cases where the case report form (CRF) allows for checking an “ongoing” or “continuing” box in lieu of providing an end date. In such cases, the SDTM-based datasets require the use of the Relative-Timing variables (i.e., --STRF, --ENRF, --STTPT, --STRTPT, --ENTPT, and --ENRTPT). When doing so, sponsors must address questions such as these: 1)”ongoing as of what point in time” (in order to ensure the correct values from the controlled terminology codelist), and 2)”is the study reference period more appropriate or is there an alternative study anchor that would be better suited?” The second area where we have seen issues is when data require the use of variables to define sample-collection time points (i.e., --TPT, TPTNUM, and ELTM), along with the anchors that identify the “reference” or baseline for these collections (i.e., --TPTREF and RFTDTC). The third area involves a misunderstanding of the differences between the --DTC and the --STDTC variables, and the associated derived variables (i.e., --DY and --STDY).

DS15 : Good vs Better SDTM: Limitations as an Operational Model
Henry Winsor, Affymax, Inc.
Mario Widel, Roche Molecular Systems
Tuesday, 2:45 PM - 3:05 PM, Location: Los Angeles/Miami

The CDISC SDTM model since 2005 has demonstrated distinct advantages to FDA reviewers when they receive data in this format. SDTM as well as ADaM are mentioned in FDA guidance as highly desirable, so more and more sponsors are including SDTM and ADaM as an integral part of the NDA/BLA/PLA submissions. While some sponsors create and submit these data sets as final products for review, others are trying to use these data sets, in particular SDTM, internally as well as in submissions. While there are distinct advantages to the latter method (fewer steps, a more efficient organization, etc.), there are many challenges that arise and need to be addressed before a sponsor can be in a position to successfully use SDTM internally. This paper addresses these challenges and offers a few proposed strategies that can be used to overcome them.

DS16 : Mapping Unique Aspects of Implantable Medical Device Study Data to CDISC SDTM Medical Device Domains
Timothy Bullock, Allergan Medical
Sini Nair, Allergan Medical
Ramkumar Krishnamurthy, Allergan Medical
Todd Gross, Allergan Medical
Tuesday, 10:15 AM - 10:35 AM, Location: Los Angeles/Miami

The Clinical Data Interchange Standard Consortium (CDISC) recently released a set of standards specifically aimed at medical device studies. These standards consist of seven proposed domains supplemental to the main Study Data Tabulation Model (SDTM). We review current methods of reporting implantable medical device study data in light of the CDISC device supplement. We report methods of modeling the general features of device-related data in the new domains. Additionally, we include a description of a custom surgery domain, SG, to capture information arising from medical devices requiring multiple concurrent, or sequential, implants as well as procedural data associated with implantation.

DS18 : ADaM Implementation Round Table
Sandra Minjoe, Accenture
Nancy Brucken, inVentiv Health Clinical
Terek Peterson , Elizabeth Li, Jeff Abolafia, Kevin Lee, Steve Kirby, Mario Widel, Karl Miller, Songlin Zhu
Monday, 10:45 AM - 11:35 AM, Location: Los Angeles/Miami

Many aspects of analysis dataset design are very study-specific. For that reason, items such as specific variables to include and the number of analysis datasets required for a given study are not defined in Version 1.0 of the ADaM Implementation Guide. Join us for an informal discussion of ADaM implementation details, and bring your own ADaM data set questions for our panel to address.

DS19 : Considerations in Data Modeling when Creating Supplemental Qualifiers Datasets in SDTM-Based Submissions
Fred Wood, Octagon Unit of Accenture
Wednesday, 9:00 AM - 9:50 AM, Location: Los Angeles/Miami

The Supplemental Qualifiers datasets described in the SDTM and SDTMIG provide a standard structure for the submission of non-standard variables (NSVs). Their naming follows the convention of SUPP--.xpt, where the two hyphens represent the two-letter domain code of the parent domain to which the NSVs belong. Despite the intended content of the SUPP-- datasets, instances have occurred where sponsors often include data in these that would actually be more accurately represented as either a separate domain or as additional records in the parent domain. Another challenge for sponsors lies in determining the merge key (represented in the IDVAR variable in SUPP-- datasets). When this is not chosen correctly, the values for the NSVs may appear on inappropriate parent records when data in the NSVs get merged onto the parent domain. In addition, cases have also been observed where sponsors have made extensive efforts to use multiple IDVAR values in the same SUPP-- dataset, when the resulting merge of the NSVs onto the parent records looks no different than if a simpler approach had been taken. This paper will present a number of actual examples where merging of the data in SUPPdatasets onto the parent record has provided some unintended and possibly erroneous data representations.

DS21-SAS : Some Strategies for Validating Your Data before Submission
Frank Roediger, SAS
Sandeep Juneja, SAS
Wednesday, 8:00 AM - 8:50 AM, Location: Los Angeles/Miami

Everyone making clinical trial submissions wants the process to go as smoothly as possible, but there are some places where missteps can cause a subsequent delay in the process. Avoiding these missteps can greatly streamline the submission preparation process. The CDISC SDTM Implementation Guides have

hundreds of pages that provide exhaustive (and exhausting) detail about every imaginable domain and variable, but those Implementation Guides don't provide any information about how long any of the variables should be. Careful data design and some utility processes can take the guesswork out of declaring variable lengths. V5 transport files are the approved mechanism for transmitting clinical trial data within a submission. Some of the archaic limitations of V5 transport files (for example, 8-character variable and format names) can now be overcome thanks to special-purpose macros that can be downloaded from SAS. The define.xml contains information about what a reviewer can expect to find in a submission. Because discrepancies between the define.xml and the submission data sets can cause delays in the FDA review, it is very important to make sure that the define.xml truly reflects the submission data's metadata. The SAS Clinical Standards Toolkit (CST) provides a way to reconcile the define.xml with submission data metadata so that any discrepancies can be resolved before a submission is made.

DS22-SAS : Assessing Drug Safety with Bayesian Hierarchical Modeling Using PROC MCMC and JMP
Doug Robinson, SAS
Monday, 9:15 AM - 10:05 AM, Location: Los Angeles/Miami

Bayesian hierarchical models are advantageous for the analysis of adverse events in clinical trials. First, the models can borrow strength across related events within the MedDRA hierarchy. Second, the models can naturally temper findings likely due to chance. We describe the implementation of two Bayesian hierarchical models (Berry & Berry, 2004; Xia et al., 2010) used for the analysis of adverse events using PROC MCMC. Once models are fit, it is necessary to review convergence diagnostics to ensure that the posterior samples of parameters sufficiently approximate the target distribution. Numerous diagnostics are available within PROC MCMC, and we also present a freely-available JMP add-in for MCMC (Markov Chain Monte Carlo) dynamically interactive diagnostics, summary statistics, and graphics.

COMPUTATIONAL SCIENCE SYMPOSIUM

CSS-WG : The PHUSE/FDA Computational Science Working Groups: A Collaboration Breaking Down the Walls Between FDA and Industry
Chris Decker, d-Wise Technologies, Inc.
Frank Senk, AstraZeneca
Mike Molter, d-Wise Technologies, Inc.
Cathy Bezek, Astellas
Tuesday, 10:15 AM - 11:35 AM, Location: Chicago Ballroom G

Over the years, walls have always stood between FDA and industry when addressing the challenges of product submissions from data standards to analysis. The goal of the PhUSE/FDA Computational Science Symposium (CSS), which was initiated last year, is to break down those walls bringing FDA and industry together around the same table to share their perspective on the issues we face and work on possible solutions and practical implementations with the goal of helping the broader community align and share experiences to advance computational science. The CSS is not only a conference but a year round collaboration where volunteers from industry and FDA work together to address these issues. This session will provide an overview of the CSS collaboration as well as a few case studies of successful projects that have been completed, such as the Study Data Reviewer's Guide, the Google code repository, and the alignment of SDTM quality checks. In addition, a panel will be available to answer questions about other projects within the initiative.

DATA VISUALIZATION AND GRAPHICS (DG)

Co-Chairs	Company
MaryAnne DePesquo	BCBS of Arizona
Nina Worden	PPD

DG01 : Managing Graphic Appearance for Grouped Data in ODS Graphics
Yunzhi Ling, Sanofi
Mei Wu, Sanofi
Tuesday, 1:15 PM - 1:35 PM, Location: Chicago Ballroom F

Analysis of clinical trial data typically involves graphical presentation of results by different treatment groups. Graphically displaying grouped data in a desired layout with specific and consistent visual attributes such as colors, marker symbols, and line patterns for each treatment group can sometimes be challenging. In ODS graphics, such challenges can be effectively overcome by properly leveraging four appearance-control components: attributes options, input data set, style template, and graph template. This paper provides some general methods and techniques for managing graphic appearance, which is often desired for grouped data in clinical research.

DG02 : Dances with Box Plot
Xuefeng Yu, Celgene
Wednesday, 10:45 AM - 11:05 AM, Location: Chicago Ballroom F

Box plot is one of the basic, commonly-used graphic tools to display the distribution of data. Making a box plot with SAS can be as easy as using a single SAS procedure, such as PROC UNIVARIATE, with specific options. It can also be complicated in some situations where the data distribution needs to be displayed in detail. This paper illustrates various SAS procedures for box plots, including the new procedures in SAS 9.2 and SAS 9.3. An advanced programming approach is discussed in detail to show how to combine a box plot with scatter plot.

DG03 : SAS GTL: Improving Patient Safety and Study Efficiency
Masaki Mihaila, Medivation Inc
Tuesday, 1:45 PM - 2:05 PM, Location: Chicago Ballroom F

Due to the high cost and time required for clinical trials, optimizing data processing is highly desirable. A graphical presentation of clinical data and relationships in Programmed Assisted Patient Narratives (PANs) is more efficient and easier to understand. Graphical tools can shorten the lengthy drug

development process while focusing attention on the review and communication of salient safety information across all company areas and functions. This paper describes how to achieve end-user friendly graphical integration into PANs by using SAS 9.3 ODS and Graphical Template Language (GTL) for 64-bit Windows. This paper only focuses on the output in RTF destination.

DG04 : A Picture is Worth 3000 words!! 3D Visualization using SAS
Suhas R. Sanjee, Novartis Institutes for Biomedical Research, INC.
Wednesday, 8:00 AM - 8:50 AM, Location: Chicago Ballroom F

Data visualization is an important aspect of clinical research. Graphs provide means to convey information efficiently in a concise manner, and three-dimensional (3D) graphs are even better, given that they provide another dimension. In general, 3D graphs seem to be underutilized in clinical research. Fortunately, SAS has a number of procedures that support data visualization in three dimensions. This paper explores these features in detail and provides examples in the oncology therapeutic area. 3D graphs are especially useful in oncology trials where the primary endpoints typically depend on a number of variables. For example, overall response computed using Response Evaluation In Solid Tumors (RECIST) criteria depends upon change from baseline, change from nadir, appearance of new lesions, etc. Plotting these variables in 3D allows for the visualization of all of these variables together. Also, a 3D surface plot can be used to visualize multiple PK concentration profiles at once.

DG05 : Customizing Survival Plots Using ODS Graphics Template Language
Fang Dong, Aastrom
Tuesday, 4:30 PM - 5:20 PM, Location: Chicago Ballroom F

SAS has made enormous strides in advancing its statistical graphics capabilities over the last several releases of the software, specifically, starting from SAS 9.1 when the ODS Graphics were experimental, to SAS 9.2 when ODS Graphics became production, and to SAS 9.3 where over 80 procedures now utilize the ODS Graphics capability. It is imperative for the programmer, including the veteran, to keep up with the advancement in order to produce highly-customized quality graphs. This paper describes several key language components necessary to produce a highly-customized graph. While the example is specific to a survival plot, the concepts should apply to customizing any other graphs.

DG06-SAS : JMP versus JMP Clinical for Interactive Visualization of Clinical Trials Data
Doug Robinson, SAS Institute
Jordan Hiller, SAS Institute
Tuesday, 2:15 PM - 3:05 PM, Location: Chicago Ballroom F

JMP software has a large set of visualization and analysis tools that can be used to analyze many different kinds of data; JMP Clinical is designed specifically for point-and-click analysis of SDTM and/or

ADaM formatted data. Both applications are widely used for cleansing, monitoring, and reviewing clinical trials data. Although JMP and JMP Clinical differ in terms of the input data format that is required and the output that is produced, they share the JMP philosophy of enabling users to produce interactive and dynamic data visualizations that enhance understanding of the data. This paper will show examples of analysis of clinical trials data in both JMP and JMP Clinical, and compare the benefits of each.

DG07 : Communication of Statistical Findings by Tables and Graphs
Howard Liang, PharmaNet/i3
Wednesday, 10:15 AM - 10:35 AM, Location: Chicago Ballroom F

When communicating clinical trial results, tables summarizing data are useful. However, if we have data summarized by graphs in addition to tables, then the presentation will be more effective. Moreover, since the summary statistics do not explain what actually happen to individual subjects, if we add some graphical displays of individual data, then we can make our conclusions more convincing. Examples will be used in this paper to show how to convert the summary statistics in a table to graphs and how to create graphs for individual subjects using the SAS/GRAPH GPLOT procedure. SAS products used include SAS/STAT: PROC MIXED, SAS/GRAPH: PROC GPLOT, and BASE SAS: PROC REPORT, PROC SQL. Operating system and SAS version: No dependency. Audience skill level: basic statistical concepts and some experience with SAS Macro programming, and using PROC GPLOT, PROC REPORT, PROC SQL and PROC MIXED

DG09 : Creating High Quality Statistical Graphs for Publications
Kriss Harris, SAS Specialists Ltd
Tuesday, 3:30 PM - 3:50 PM, Location: Chicago Ballroom F

Do you want to produce high quality graphs for publications or presentations? Do you want to add p-values on the graphs with annotations that show which groups are being compared? Do you want to learn more about the DPI option and how to use it? If you answered yes to any of the questions or have an interest in SAS Graphics, then this paper is for you. This paper will demonstrate how to do the above plus show how to create first-rate Kaplan Meier Graphs, Forest Plots and Napoleon Plots using SAS 9.2.

DG10 : Data Visualization Tips and Techniques for Effective Communication
LeRoy Bessler, Bessler Consulting and Research
Wednesday, 9:00 AM - 9:50 AM, Location: Chicago Ballroom F

This tutorial presents my principles of communication-effective data visualization, and shows widely usable ways to implement them with new technology (ODS Graphics and Statistical Graphics Procedures), or, when the best choice, old technology (SAS/GRAPH Procedures). Most of the solutions

use the new technology, which is imbedded in Version 9.3 of Base SAS software at no additional cost. I will identify what the new technology can do best, and what to be wary of when using that technology.

DG11 : Data Visualization Power Tools: Expedite the Easy, Implement the Difficult, or Handle Big Data
LeRoy Bessler, Bessler Consulting and Research
Tuesday, 4:00 PM - 4:20 PM, Location: Chicago Ballroom F

The power tools are SAS macros for ODS Graphics or SAS/GRAPH. You can use them as is, modify them for similar, but different, tasks, or learn from them to build new solutions for very different problems. Power tools enable you to create results quickly and easily. Power tools that are poorly designed enable you to create scrap quickly and easily. Here, the standard of good design is communication effectiveness of the result. Some of the tools presented can enable you to show what's important when confronted with too much information, and to fit whatever is of importance into the web browser window, preventing any need for scrolling. To interpret a picture, you need to see the whole picture at one time. Other tools presented, which are not web-specific, but can be used in that context, enable you to get the best result out of SAS graphic software with minimal coding.

HANDS-ON TRAINING (HT)

Co-Chairs	Company
Raoul Bernal	Genentech
Alissa Ruelle	inVentiv Health Clinical

HT01 : Quick Results with PROC SQL
Kirk Paul Lafler, Software Intelligence Corporation
Monday, 9:15 AM - 10:35 AM, Location: Denver/Houston/Kansas City

Structured Query Language (SQL) is a universal language that allows you to access data stored in relational databases or tables. This hands-on workshop presents core concepts and features on using PROC SQL to access data stored in relational database tables. Attendees learn how to define, access, and manipulate data from one or more tables using PROC SQL quickly and easily. Numerous code examples are presented on how to construct simple queries, subset data, produce simple and effective output, join two tables, summarize data with summary functions, create and use "virtual" tables, and create and use indexes.

HT02 : Using PROC FCMP to the Fullest: Getting Started and Doing More
Arthur Carpenter, CALOXY
Tuesday, 8:00 AM - 9:20 AM, Location: Denver/Houston/Kansas City

The FCMP procedure is used to create user-defined functions. Many users have yet to tackle this fairly new procedure, while others have only attempted to use only its simplest options. Like many tools within SAS®, the true value of this procedure is only appreciated after the user has started to learn and use it. The basics can quickly be mastered and this allows the user to move forward to explore some of the more interesting and powerful aspects of the FCMP procedure. Starting with the basics of the FCMP procedure, this paper also discusses how to store, retrieve, and use user-defined compiled functions. Included is the use of these functions with the macro language as well as with user-defined formats. The use of PROC FCMP should not be limited to the advanced SAS user; even those fairly new to SAS should be able to appreciate the value of user-defined functions.

HT03 : Hands-On ADaM ADAE Development
Sandra Minjoe, Octagon Research Solutions
Tuesday, 10:15 AM - 11:35 AM, Location: Denver/Houston/Kansas City

The Analysis Data Model (ADaM) Data Structure for Adverse Event Analysis was released by the Clinical Data Interchange Standards Consortium (CDISC) ADaM team in May, 2012. This document is an appendix to the ADaM Implementation Guide (IG), and describes the standard structure of the analysis dataset used for most of our typical adverse event reporting needs. This hands-on training focuses on creating metadata for a typical adverse event (AE) dataset. Attendees will work with sample SDTM and ADaM data, finding information needed to create the results specified in a sample set of table mock-ups. Variable specifications, including coding algorithms, will be written. Some familiarity with SDTM data, AE reporting needs, SAS DATA step programming, and Microsoft Excel is expected. Attendees will also learn how to apply the data structure for similar analyses other than adverse events.

HT04 : Effectively Utilizing Loops and Arrays in the DATA Step
Arthur Li, City of Hope
Wednesday, 8:00 AM - 9:20 AM, Location: Denver/Houston/Kansas City

The implicit loop refers to the DATA step repetitively reading data and creating observations, one at a time. The explicit loop, which utilizes the iterative DO, DO WHILE, or DO UNTIL statements, is used to repetitively execute certain SAS statements within each iteration of the DATA step execution. Utilizing explicit loops is often used to simulate data and to perform a certain computation repetitively. However, when an explicit loop is used along with array processing, the applications are extended widely, which includes transposing data, performing computations across variables, etc. Being able to write a successful program that uses loops and arrays, one needs to know the contents of the program data vector (PDV) during the DATA step execution, which is the fundamental concept of DATA step programming. This workshop will cover the basic concepts of the PDV, which is often ignored by novice programmers, and then will illustrate utilizing loops and arrays to transform lengthy code into more efficient programs.

HT05 : The Armchair Quarterback: Writing SAS® Code for the Perfect Pivot (Table, That Is)
Peter Eberhardt, Fernwood Consulting Group Inc.
Monday, 1:15 PM - 2:35 PM, Location: Denver/Houston/Kansas City

“Can I have that in Excel?" This is a request that makes many of us shudder. Now your boss has discovered Excel pivot tables. Unfortunately, he has not discovered how to make them. So you get to extract the data, massage the data, put the data into Excel, and then spend hours rebuilding pivot tables every time the corporate data are refreshed. In this workshop, you learn to be the armchair quarterback and build pivot tables without leaving the comfort of your SAS environment. In this workshop, you learn the basics of Excel pivot tables and, through a series of exercises, you learn how to augment basic pivot tables first in Excel, and then using SAS. No prior knowledge of Excel pivot tables is required.

HT06-SAS : Using the SAS Clinical Standards Toolkit 1.5 to Import CDISC ODM files
Lex Jansen, SAS Institute Inc.
Tuesday, 3:30 PM - 4:50 PM, Location: Denver/Houston/Kansas City

The CDISC Operational Data Model (ODM) is a vendor-neutral, platform-independent XML format for interchange and archival of clinical study data. The model represents study metadata, administrative metadata, reference data and subject data associated with a clinical trial. The ODM format is defined by an XML schema and a specification. The SAS Clinical Standards Toolkit (CST) is a framework that supports ODM. To enable this support, SAS has defined a relational data model that represents the ODM model as SAS data sets. This Hands-on Training will provide an introduction to the structure and content of ODM files and shows how CST supports: Importing ODM files, schema-level validation of an ODM file, validating structure and content of the SAS representation of ODM, extraction of clinical data and reference data into SDTM SAS datasets, and importing CDISC NCI Controlled Terminology ODM files.

HT07-SAS : Some Techniques for Integrating SAS Output with Microsoft Excel Using Base SAS®
Vince DelGobbo, SAS
Monday, 3:30 PM - 4:50 PM, Location: Denver/Houston/Kansas City

This paper explains some techniques to integrate your SAS output with Microsoft Excel. The techniques that are presented in this paper require only Base SAS 9 software, and can be used regardless of the platform on which SAS software is installed. You can even use them on a mainframe! Creating and delivering your workbooks on-demand and in real time using SAS server technology is discussed. Although the title is similar to previous papers by this author, this paper contains new and revised material not previously presented.

HT08 : SDTM, ADaM and define.xml with OpenCDISC
Angela Ringelberg, inVentiv Health Clinical
Tracy Sherman, inVentiv Health Clinical
Tuesday, 1:15 PM - 2:35 PM, Location: Denver/Houston/Kansas City

As programmers, many of us have spent hours reviewing SDTM/ADaM standards and implementation guides to generate “compliant” CDISC SAS data sets. Is there an easier way to ensure compliance with CDISC standards, including SDTM, ADaM, SEND, Define.xml, and others? OpenCDISC is an open source community which is focusing on creating frameworks and tools for the implementation and advancement of CDISC Standards. OpenCDISC has created a CDISC Validator, which will eliminate the need for individuals to develop their own custom processes in order to ensure that their CDISC models are compliant with CDISC standards. By taking common validation rules, OpenCDISC has developed an open-source tool which is freely available and of commercial quality to ensure data compliance with CDISC models such as SDTM, ADaM, SEND and Define.xml. The validation rules for each standard have been pooled into a CDISC Validation Rules Repository, providing users with a central listing. The listing is

easy to use, modify and continue development. In this Hands-On Training, we are going to briefly describe a few of the key terms (SDTM, ADaM, Define.xml) and investigate the use of OpenCDISC Validator to perform the validation of SDTM 3.1.1 and 3.1.2 SAS data sets, ADaM 1.0 SAS data sets and define.xml. We will also show you how to generate a shell for your define.xml.

HEALTH OUTCOMES AND EPIDEMIOLOGY (HO)

Co-Chairs	Company
Richard Allen	Peak Statistical Services
Matt Karafa	Cleveland Clinic

HO01 : Obesity and Weight Cycling Use SAS Software for Epidemiological Studies
Marina Komaroff, Noven Pharmaceuticals
Monday, 9:15 AM - 9:35 AM, Location: Chicago Ballroom G

The epidemic of obesity is a serious problem. According to CDC statistics for the United States population, about 35.7% of adults and approximately 17% of children and adolescents (2-19 years old) were obese in 2009-2010. Numerous epidemiological studies demonstrated that obesity is a risk factor for multiple diseases like cancer and cardiovascular. Various weight loss and weight control programs were conducted in recent years and demonstrated effectiveness for a short period of time; yet, the majority of participants regained weight sooner or later. The process of losing and regaining weight is called weight cycling, and the prevalence of weight cyclers is growing in the United States. The phenomenon of weight cycling is less studied. Some research for the association of weight cycling and morbidity and mortality was conducted in the 1990's and demonstrated controversial results. The National Task Force on the Prevention and Treatment of Obesity summarized the forty three (43) English-language articles that evaluated the effects of weight cycling on humans or animals in studies done from 1966 through 1994. The authors concluded that most studies demonstrated the association between body weight cycling and mortality and morbidity. Nevertheless, the biggest problem in deriving solid conclusions was the lack of a standardized definition of weight cycling, and research slowed down in this area. This paper provides a brief review of proposed methodological approaches for weight cycling. The most critical key elements of weight cycling like the magnitude of gain/loss, duration and the frequencies-number of cycles are suggested for a unified algorithm. A user-friendly SAS V9.1.2 macro that is flexible in choosing parameters and transparent in each step of calculation identifies weight cyclers. The author is convinced that this paper and proposed algorithm supported by a SAS macro can expedite the research on the harm or benefit of weight cycling on the health of millions of people.

HO02 : Practice of SMQs for Adverse Events in Analysis of Safety Data and Pharmacovigilance
Gary Chen, Shire Pharmaceuticals
David Shen, Independent Consultant
Monday, 9:45 AM - 10:05 AM, Location: Chicago Ballroom G

MedDRA cannot group adverse events that indicate the presence of a medical condition through PTs properly due to their high granularity. SMQs were developed specifically to address this issue and to maximize the likelihood that all terms related to a specific medical condition of interest can be identified. This paper describes the practical implementation of SMQs and how SMQs can be applied to search adverse events for statistical analysis and create tables, listings and graphics. SAS and PROC SQL are used to read in the SMQ dictionary files and then convert into appropriate data structures in searching for adverse events of concern. The practical techniques we present offer a good overview of SMQs and their applications in analysis of safety data and pharmacovigilance. This paper will educate all levels of SAS users who are interested in safety data analysis and pharmacovigilance with SAS programming.

HO03 : Imputing Dose Levels for Adverse Events
John R Gerlach, CSG, Inc.
Igor Kolodezh, CSG, Inc.
Monday, 10:15 AM - 10:35 AM, Location: Chicago Ballroom G

Besides the standard reporting of adverse events in clinical trials, there is a growing interest in producing similar analyses in the context of exposure to treatment drug at the onset of an adverse event. Given an ADaM data set containing adverse events (ADAE), the intended analysis initially requires the inclusion of a variable denoting the dose level at the onset of an adverse event, called DOSEAEON. This variable would contain a null value for non-treatment emergent events, zero for placebo, and indicate the dose level, otherwise. Moreover, DOSEAEON would be used to create grouping variables for the actual analysis. This paper discusses the challenges of implementing a hierarchical methodology to determine the dose level at the onset of an adverse event.

HO04 : MedDRA Beyond That Basic AE Report: How SMQs and MedDRA Structure Can Enhance Reporting
Pamela Giese, inVentiv Health Clinical
Monday, 10:45 AM - 11:05 AM, Location: Chicago Ballroom G

The basic Adverse Events (AE) table, organized by MedDRA System Organ Class (SOC) is common across clinical reporting. Yet there is a whole structure of MedDRA that allows for more targeted reporting. In addition, MedDRA includes Standard Medical Queries (SMQs) whereby events (or histories) can be summarized across SOCs. SMQs can be especially helpful in pharmacovigilance, late phase, and health

outcome analysis. This paper will describe the basic structure of MedDRA, the relationship of SMQs and SOCs, and examples of how these can be helpful.

HO05 : Taking a Census in Utero: An Introduction to Pregnancy Registries with an Emphasis on Identifying Multiple Gestations

Britney Gilbert, InVentiv Health Clinical

Monday, 11:15 AM - 11:35 AM, Location: Chicago Ballroom G

Clinical trials cannot effectively assess drug therapies and their effects within pregnant subjects, because of the ethical questions surrounding experimental treatments and their risks to the mother and fetus. Therefore, pregnancy registries are established for safety surveillance of exposures before and after conception. In these registries, analysis involving outcome statistics must go beyond simple demographical descriptions and disclose all confounding biases to properly understand the population. According to the National Vital Statistics Report, the twinning rate has increased more than 70% from 1980 to 2009 and the rate of higher order multiple births (triplet/+) increased more than 400% during the 1980s and 1990s. With these rate increases, it becomes important to properly identify multiple gestations within a pregnancy registry since they will impact the analysis, in particular the spontaneous abortion rates, preterm birth rates, and low birth weight rates. The purpose of this paper is to introduce programmers to pregnancy registries and raise awareness of how multiple gestations affect outcome analyses.

INDUSTRY BASICS (IB)

Co-Chairs	Company
Sandy Patternotte	PPD
Mario Widel	Roche

IB01 : Serving SAS®: A Visual Guide to SAS Servers
Greg Nelson, ThotWave
Tuesday, 3:30 PM - 4:20 PM, Location: Chicago Ballroom C

SAS has been running on servers since the late 1960s. Despite the emergence of PCs and workstation-class machines, SAS still reigns supreme on the server. With the introduction of the SAS platform 9 in 2004, the number and types of servers have grown exponentially. As any good student of the DATA step will attest, knowing what SAS is doing is a critically important step in debugging and authoring efficient programs. In this paper, you will experience SAS through a visual tour - you will see what SAS is doing, how it works, which server is doing what, when the operating system plays a role, how security functions, and what happens to your data through the entire process.

IB02 : Analysis of Concomitant Medication Data
JJ Hantsch, inVentiv Health Clinical
Karl Miller, inVentiv Health Clinical
Wednesday, 10:45 AM - 11:05 AM, Location: Chicago Ballroom C

Concomitant medication data (in parlance, conmeds) is mandated by GCP for clinical trials. However, conmeds are frequently unanalyzed. Well-analyzed conmeds can indicate which adverse events and first-line medications are linked to the experimental medication, and which medications may interact. The integrated summary of safety (ISS) provides an opportunity to perform a good conmed analysis before the approval process is complete and in time for label writing. Any conmed analysis must consider all experimental subjects exposed to both experimental medication and the placebo. Analyzing the placebo-exposed subjects provides information about a baseline level of concomitant medications. Any category of medications which exceeds a pre-determined threshold, often 5%, is reported. In CDISC, conmed data is reported in the CM domain. Analysis can be accomplished straight from this raw data or when merged with MH, AE, EX, SU, DM or ADSL data into an ADaM dataset. Adverse events are reported and grouped with the contemporary version of MedDRA. Conmeds are normally reported and grouped by ATC code level. The ideal conmed analysis would identify 1) all medications (or medication classes) which are reported for greater than a threshold level (5%), 2) all medications where the difference between placebo and experimental is greater than a threshold level (1%) 3) all medications with a significant difference in the distribution of adverse events between the two groups and 4) all medications which differ by a threshold level (10%) between significant outcome categories (e.g. early withdraws, SAE, successful treatment.)

IB03 : Due (Data) Diligence: Study Data Review for Acquisitions
Steve Kirby, ViroPharma, Incorporated
Adam Young, ViroPharma, Incorporated
Tuesday, 4:30 PM - 4:50 PM, Location: Chicago Ballroom C

Drug companies have increasingly used acquisitions to leverage promising drug candidates. Ensuring that the price paid for a drug candidate is supported by clinical results naturally demands thorough review prior to purchase, but drilling down to investigate whether those results can be easily traced back to the clinical data can be neglected until close to the time of submission to a regulatory body. As companies work to identify and purchase promising drugs, available study data should be carefully evaluated for consistency with regulatory submission requirements. Small deviations from submission requirements for study data can have a dramatic effect on timelines. If those issues are identified early in the acquisitions process, mitigation plans can be put into place to minimize downstream impact and price can be adjusted to reflect the added time and effort needed to get the drug to market. Starting with the best case scenario, results supported by complete SDTM and ADaM data and comprehensive documentation, the presenters will explore ways to determine whether any deviations from that best case scenario exist and will suggest how to quantify the effect that those deviations could have on a purchasers ability to submit the drug for approval.

IB04 : Map Metadata Taking the Next Step/Connecting the Dots: How Metadata Can Go Beyond Data Standards to Define Data Movement
Gregory Steffens, Novartis
Praveen Garg, ICON Development Solutions
Tuesday, 1:15 PM - 2:05 PM, Location: Chicago Ballroom C

Much attention has been paid to the design and use of metadata to store standards, study data specifications and creating the define.xml. But not as much attention has been focused on the design and use of metadata to transform and derive data. We must move beyond to focus on the stopping points of data flow and concentrate on the data movement from one stopping point to another. This paper describes the need for an industry standard for map metadata and presents a design that has been implemented with great success. Just like metadata, map metadata should not assume any one standard. Well-designed map metadata supports meta-programming of data flow and not just implementing database attributes. Map metadata has many other advantages, such as data flow transparency, internally and during regulatory submissions, and standardized ways to exchange specifications about how data flows from one structure to another, as in SDTM to ADaM to IDB. Map metadata defines the relationship between the source and target database at data set, variables, row and value level. Map metadata along with source and target metadata can help to automate the data flow and create transparent define files (with transformation logic as well as description). Once the standard map metadata and standard target metadata is defined, one can create the study-level metadata based on source metadata. This is the next evolution of metadata and of meta-programming, leading to a true Data Transformation Engine (DTE).

IB05 : The Value of an Advanced Degree in Statistics as a Clinical Statistical SAS Programmer
Mark Matthews, inVentiv Health Clinical
Ying (Evelyn) Guo, Parexel
Wednesday, 8:30 AM - 8:50 AM, Location: Chicago Ballroom C

Clinical statistical programmers often leverage the SAS software to process data in the form of a data set, table, figure or listing. Frequently there are a set of rules, also known as programming specifications, which can enable a non-statistician to compute some of the most complex statistical results. These specifications are generally created by a seasoned biostatistician. However, with current industry trends, more advanced statistical methods are available. The statistician needs to be free from creating those programming specification and rules to explore better analysis methods. A SAS programmer with an advanced degree in statistics is quite capable of closing those gaps which makes the SAS programmer, who is a statistician, a very effective role in the clinical field.

IB06 : Introduction to the CDISC Standards
Sandra Minjoe, Octagon Research Solutions
Wednesday, 9:00 AM - 9:50 AM, Location: Chicago Ballroom C

The Clinical Data Interchange Standards Consortium (CDISC) encompasses a suite of standards across the clinical space. The Study Data Tabulation Model (SDTM) and Analysis Data Model (ADaM) are probably the two standards most familiar to PharmaSUG attendees, but there are many others. This paper and presentation focus on the foundational standards of CDISC, from protocol to analysis reporting, along with data exchange and controlled terminology. Each of these standards is introduced, emphasizing how it fits into the big picture. Whether you want to discover these standards that are quickly becoming common place in our industry, or just add a few acronyms (PRM, CDASH, SDTM, SEND, ADaM, ODM, and NCI EVS) to your vocabulary, then this is for you!

IB07 : Clinical Trials: If You Can Explain Them to Second Graders, You Can Explain Them to Anyone!
Lara Guttadauro, inVentiv Health Clinical
Tuesday, 5:00 PM - 5:20 PM, Location: Chicago Ballroom C

We work in a technical field and face complicated challenges daily. Discussing these issues with like-minded colleagues can be difficult at times. Another layer of complexity is added when you need to explain an issue to professionals outside your department. Yet another layer is added when you try explaining your career to friends and family. Now imagine explaining clinical trials to second graders! This is the challenge I faced when asked to participate in career day at my daughter's school. In this paper, I share my career day experience with hopes that you can benefit the next time you have a challenging issue to explain to a diverse audience.

IB08 : Pretty Please?! Making RTF Output "Pretty" with SAS
Carol Matthews, United Biosource Corporation
Elena Kalchenko, United Biosource Corporation
Tuesday, 2:15 PM - 3:05 PM, Location: Chicago Ballroom C

Output generated from SAS in the Pharmaceutical Industry invariably ends up in a document that is presented to regulatory authorities, review boards, publications, and a variety of other audiences. As SAS has evolved to more easily generate output that can be read directly into software packages such as Microsoft Word, the expectation that SAS output will look “published” is becoming the standard. Recipients of SAS output now expect actual superscript “a” as a footnote symbol rather than a regular text “a” or ”*” symbol. While some features such as fonts, margins and tabs can be controlled directly with SAS, other features need to be added with RTF code. In this paper, we will discuss the combination of SAS and RTF code that is required to generate output directly from SAS that looks like it was word-processed. We will touch on the basic components of PROC TEMPLATE needed to control the margins and overall appearance of output, the basic ODS syntax needed to control where output goes and how SAS interacts with RTF, various features of PROC REPORT that can be used to further refine the look of output, and common RTF code that can be used to put the final polish on tables and listings.

IB09 : Is Your Data Set Really Validated: Beware of “Blind-Fold” Validation
Neha Mohan, inVentiv Health Clinical
Gayatri Karkera, inVentiv Health Clinical
Wednesday, 10:15 AM - 10:35 AM, Location: Chicago Ballroom C

The aim of validation programming in clinical trials is to produce documented evidence that the trial data and reports meet the quality attributes as per clinical requirements. One of the basic and minimal necessities during validation is to match the numbers in both of the outputs, original as well as validation. In order to achieve this, some validation programmers often inadvertently perform what we term as “blind-fold” validation in this paper. We call it “blind-fold” validation because many times the programmers have a tendency to follow the specifications “as is” - down to the punctuation, and modify the validation output to match the original output if any differences are noticed. This paper illustrates with simple examples of data set validation from clinical trials how these practices can inadvertently lead to erroneous outputs passing validation. It then offers logical and meaningful validation approaches that hold valid beyond data sets as well to help improve quality and reliability of the data using the get it right the first time principle.

IB10 : SAS Programmer to Clinical SAS Programmer
Gayatri Karkera, inVentiv Health Clinical
Neha Mohan, inVentiv Health Clinical
Wednesday, 8:00 AM - 8:20 AM, Location: Chicago Ballroom C

In the clinical programming industry, as hiring managers, we tend to select candidates with varied academic backgrounds spanning from life sciences to abstract subjects like mathematics, statistics and computer sciences. Many times our selection bias for candidates with mathematics, statistics and computer sciences, as compared to candidates from other academic backgrounds, eminates purely from their training in programming skills. Anyone who has been in the clinical programming industry for long would agree that SAS programming in this industry requires a lot more than just programming. In this note, we attempt to bring out some dimensions of the journey from pure SAS programmer to clinical SAS programmer. The trajectory of this exciting journey is full of learning. We share some of our learning and experiences based on our industry observations.

MANAGEMENT AND SUPPORT (MS)

Co-Chairs	Company
Elaine Dempsey	University of North Carolina
John Labore	Eli Lilly and Co.

MS01 : Modern SAS Programming: Using SAS Grid Manager and Enterprise Guide in a Global Pharmaceutical Environment
David Edwards, Amgen
Greg Nelson, ThotWave
Monday, 10:15 AM - 10:35 AM, Location: Chicago Ballroom H

Amgen, like most large biotechnology companies, uses SAS to support the drug discovery process. Used throughout the organization for data management, analytics and reporting activities, SAS use extends to research, operational and manufacturing units in the organization. Like many organizations, Amgen has grown organically and their use of SAS is no exception. Equipped with a vision to fully leverage its global workforce and to maximize their IT investments, Amgen developed a SAS-based, research informatics infrastructure to deliver value around the globe. This paper will highlight many of the aspects of this project including business justification, requirements, design, verification and validation and production migration for over 1500 programmers and statisticians spread across three continents. We will highlight some of the challenges we faced and how these were overcome using improved processes, modern technologies such as SAS Grid Manager and Enterprise Guide and the combined efforts of a global project team. Having just finished the third and final wave of production migration, we will be able to talk in detail about what worked and what could have been improved both from a project perspective as well as the technologies deployed.

MS02 : Life Cycle of a Data Point - A Tool to Educate the Clinical Development Team on What it Takes to Go From the Clinic to a Submission
Arthur Collins, Biogen Idec, Inc.
Joanna Koft, Biogen Idec, Inc.
Monday, 9:45 AM - 10:05 AM, Location: Chicago Ballroom H

A common issue facing Biometrics and Data Management groups is a lack of understanding from other members of the Clinical Development Team, leading to unrealistic expectations and pressure to shorten timelines. A cross-functional group from Biostatistics, Statistical Programming, Statistical Submissions Management, Data Standards, and Data Management got together and developed a presentation to show all of the steps that must be carried out to take data from the point of collection to being included in a submission and discuss how deviating from standards can impact each step. This presentation was

very well-received by a group of MDs from Clinical and Drug Safety and was recorded as a standalone presentation to be included as mandatory training for those groups and for Clinical Operations going forward. This paper describes the process of developing the presentation and the contents. The conference presentation will include excerpts from the recorded presentation as well.

MS03 : Using Workflows and Metadata Information to Standardize Business Processes in Pharmaceutical Programming
Peng Yang, Santen
Wei Liu, Santen
Julie Maddox, SAS Institute
Monday, 2:15 PM - 2:35 PM, Location: Chicago Ballroom H

Pharmaceutical programming teams are getting increasingly more distributed and the tasks are becoming more complex. Standardizing business processes and ensuring consistency is crucial to the successful statistical analysis of clinical studies. Workflow templates can be defined to combine business process activities with the best practice techniques to be used as standards across an organization. Workflows can integrate both manual steps and complex processing steps that can use programmatic actions. This paper describes how workflows can enable several common business processes in clinical data management, biostatistics and statistical programming. In addition, it demonstrates in technical detail how the SAS Drug Development system implements these workflows. Possible application areas of workflows in the pharmaceutical programming will be discussed, such as workflows governing the development and validation of statistical programs using independent double programming practice, the interaction between statisticians and programming teams, and workflows for management of standards data exchange between a sponsor and the external partner (CRO). SAS Drug Development also provides users with API macros, which can be used to extract metadata information from SAS programs, logs and outputs. This enables the lead programmers or programming managers to visualize and manage many different aspects of the statistical programming projects. Finally, the workflow management system and metadata capability in SAS® Drug Development contributes to optimizing the described business processes and allowing management to make fact-based decisions, improve quality and efficiency of registrational filings, and optimize the conduct of future clinical trials.

MS04 : Deployment of SAS® Programming Contract Staff: Pathway to Sinkhole or Best Strategy for Programming Division?
Parag Shiralkar, Eliassen Group
Monday, 10:45 AM - 11:05 AM, Location: Chicago Ballroom H

In the pharmaceutical industry, various factors decide the budget associated with SAS programming services. These factors often relate to hiring policies, roles and responsibilities defined within the organization, and strategy defined by upper management as far as utilization of budgets is concerned. Traditionally, SAS programming is often considered as an “allied” function supporting other main

functions like statistics and data management. Considering these circumstances and substantial fluctuations in resourcing needs, the programming division is often “forced” to adopt a strategy of hiring contract staff. Contract staff deployment in SAS programming often provides lot of benefits to the programming division. Such benefits are realized through co-employment risk reduction and no obligation for the division to keep the augmented contract staff on-board when project is over. At the same time, this strategy may backfire on the division if issues arise from recruitment, training, and retention of contract staff. If issues arise, they may jeopardize the ability of the programming division to complete their projects or meet project needs, and may risk loss of credibility as well. This manuscript sheds light on the core aspects of contract staffing, and discusses the risks and benefits of contract staffing of SAS programmers in the pharmaceutical industry, especially supporting clinical trials. Through case studies, this manuscript provides more informative discussion about things which procurement and programming departments should pay attention to while defining their strategy of hiring contract staff.

MS05 : Be a Dead Cert for a SAS Cert: How to Prepare for the Most Important SAS Certifications in the Pharmaceutical Industry
Hannes Engberg Raeder, inVentiv Health Clinical
Monday, 11:15 AM - 11:35 AM, Location: Chicago Ballroom H

It is a fact that SAS has become the standard for analyzing and reporting clinical data in the pharmaceutical industry; thus, it is one of the most important tools for the clinical database programmer, statistical programmer and statistician. Getting certified raises your overall programming awareness, improves the quality of your work and earns official recognition from SAS. There are currently three certifications of relevance for the clinical SAS programmer: SAS Certified Base Programmer, SAS Certified Advanced Programmer and SAS Certified Clinical Trials Programmer. However, lack of time and good learning strategies are potential obstacles that may discourage candidates from trying. Based on my own experience, the aim of this paper is to provide an overview, resources for self-study, and advice on dealing with practical aspects like integrating learning into daily work and on the exams themselves, such as the type of questions encountered and exam strategy.

MS06 : Managing 21 CFR Part 11 Compliance: Using Checksums On Open Systems
Carey Smoak, Roche Molecular Systems, Inc.
Mario Widel, Roche Molecular Systems, Inc.
Monday, 2:45 PM - 3:05 PM, Location: Chicago Ballroom H

Generally, closed systems are encouraged in order to ensure compliance with 21 CFR Part 11 regulations. Databases with audit trails are typically considered to be examples of closed systems, which are designed to ensure the authenticity, integrity, and electronic records confidentiality while the signer cannot repudiate the signed record as not genuine. Open systems do not have these capabilities. In this paper the authors will show how to address record authenticity and integrity using checksums. They will describe the managing of electronic lab instrument data in an open system. Checksums can be used to

ensure that data has not been accidentally modified during the process of acquisition from the lab instrument to its final destination as a SAS data set on a secure server. While the lab instrument data is not maintained in a database with an audit trail, it is shown how the spirit of 21CFR Part 11 compliance is kept with respect to data integrity and record authenticity.

MS07 : If 'Standard Code' is so Great…
Brian Fairfield-Carter, ICON Clinical Research
Monday, 9:15 AM - 9:35 AM, Location: Chicago Ballroom H

Since CROs traditionally employ 'lean', 'do more with less' project teams, their use is increasingly attractive to sponsors as a cost-trimming solution. In the analysis and reporting world, the push for ever-greater efficiency typically focuses on the concepts of standardization and code re-use, and specifically on 'standard macro libraries'. But efficiency is an elusive concept: anyone who has performed code maintenance knows that gains in processor time can easily be dwarfed by issues of 'human readability', and that modifying code can turn out to be more time-consuming and error-prone than writing code from scratch; standard macros can be restrictive and involve a considerable learning curve. A review of conference papers dealing with 'standard code' illustrates a consistent thread: standard macro libraries, at least those involved in analysis and reporting, exist almost exclusively in the domain of pharmaceutical companies, and are virtually non-existent in CROs. This begs the question: if 'standard code' works in certain contexts, and seems intuitively to offer efficiency gains, why does it not seem to hold greater prominence in the CRO world? This paper seeks to catalog the trade-offs inherent to code libraries, and to challenge the conventional dogma that standardization inevitably produces efficiency/accuracy gains, in exploring the near absence of macro libraries in the CRO world. This discussion then provides a backdrop for proposing what may be a more workable approach to code re-use: one that focuses on planned code adaptation, and on code organization and architecture, where code writing deliberately anticipates future adaptation.

MS08 : Therapy Lessons Learned to Empower Programmers
Shelley Dunn, d-Wise Technologies, Inc..
Monday, 4:30 PM - 4:50 PM, Location: Chicago Ballroom H

In the same way that educators teach much more than their subject matter, managers of programmers teach their employees much more than programming. Essentially teachers teach students and managers teach their employees. While it is imperative to know your subject matter in order to teach it to students, or to know how to program in order to manage programmers, quite often, the interpersonal relationship that a manager maintains with her or her employees will have much more influence on an employee’s work quality, productivity, job knowledge, creativity, job satisfaction, ability to work with others, and a whole host of other criteria which often form the basis of performance evaluations. So how does this relate to therapy? First of all, this is not to imply that all managers or all programmers need therapy. Nor is it meant to imply that those who have been in therapy somehow have a leg up on

those who haven't. However, many of the same communication techniques learned and used in a therapeutic environment between a therapist and patient can be used in an office setting to promote a collaborative and supportive environment in which programmers can feel empowered. This paper is geared towards those who manage programmers. However, much of the information can also be applied to those who manage people in general.

MS09 : The Challenges and Opportunities for SAS Statistical Programmers in Two Commonly used CRO Resourcing Models
Mark Matthews, inVentiv Health Clinical
R. Mouly Satyavarapu, inVentiv Health Clinical
Monday, 1:15 PM - 1:35 PM, Location: Chicago Ballroom H

Managing one's own career and aspirations as SAS programmers in the clinical trial space is always of interest. The Contract Research Organization (CRO) industry provides a diverse opportunity for SAS programmers to pursue and experience their daily work satisfaction and career. Two commonly used models in the CRO space are traditional "Deliverable Based Model" and "Full-Time Equivalent (FTE) Time and Material Model" (also called as role based model). Previously, a presentation at PharmaSUG 2012 identified and differentiated various parameters and methods followed by each model (See PharmaSUG 2012 - Satyavarapu). This paper will take some of those parameters and share actual experiences on how they can provide influence on the career direction of a CRO SAS statistical programmer. The discussion will extend into 3 main categories: SAS and technical skills development, leadership and interpersonal skills development, and the hiring and working conditions. This can give the SAS programmer a better idea on what to expect in these two working environments. Both environments offer challenges and opportunities that can be considered when choosing the best fit for your interests and career.

MS10 : Transition from "Hands-On" Statistical Programmer to Leader of a Team of Role-Based Statistical Programmers: Tools and Tips.
Rodrigo Juarez y Ruiz, Eli Lilly Canada Inc.
Monday, 5:00 PM - 5:20 PM, Location: Chicago Ballroom H

This paper is inspired by Jim Grudzinski's PharmaSUG 2012 presentation titled "Managing a Blended Programming Staff of Permanent Employees and Contingent Workers". However, it also shares my personal experiences and perspectives as team leader (i.e. the permanent employee), being responsible for managing projects and coordinating workers in a "role based" sourcing model (i.e. the contingent worker). I will describe in detail what I consider four basic steps: attitude, knowing your team, adequate prioritization and tracking, as well as recommendations, tips, and my opinions in leading and coordinating role-based statistical programmers that have helped me successfully execute and finish projects that were both planned and unplanned, simple and complex. This presentation is not meant to be a seminar on project management nor does it represent the views and policies of Eli Lilly or its

affiliates; it rather reflects my personal opinion on how to successfully transition from permanent staff doing “hands-on”/”heads-down” work to doing project and resource management while leading a team of role-based statistical programmers.

MS12 : Statistical Computing Environment Implementation - An Agile Approach
Gary Cozzolino, d-Wise Technologies
Monday, 1:45 PM - 2:05 PM, Location: Chicago Ballroom H

Historically, companies in the pharmaceutical industry have stored SAS data, programs, and output on Windows or UNIX file systems, and executed SAS on shared servers or locally on PCs. More organizations are now moving to “Statistical Computing Environments” (SCEs), which provide secure, audit-trailed computing platforms, enabling programmers to effectively manage their programs through the Software Development Lifecycle. As more companies move to SCEs, pharmaceutical, biotech, and CRO organizations face a number of challenges in implementing such systems. Regardless of the technology involved, and whether a commercial product is used or if it’s a home grown system, significant changes to related business processes should be expected. Among these challenges, programmers will need to learn new processes for program development, management and execution. This presentation will outline an agile approach to implementing these enterprise solutions. The iterative methodology begins with identification of priority business processes, followed by rapid piloting and revisions, leading to development of formal processes for production. Other key aspects to be considered include change management, training, and study migration to the new environment.

MS14-SAS : Patient Profile Graphs using SAS
Sanjay Matange, SAS
Monday, 3:30 PM - 4:20 PM, Location: Chicago Ballroom H

Patient profiles provide information for a specific subject participating in a study. The report includes relevant data for a subject that can help in correlating adverse events to concomitant medications and other significant events as a narrative or a visual report. This presentation will cover the creation of the graphs useful for visual reports from CDISC data. This includes a graph of the adverse events by time and severity, graphs of concomitant medications, vital signs, and labs. All the graphs are plotted on a uniform timeline, so the adverse events can be correlated correctly with the concomitant medications, vital signs and labs. These graphs can be easily incorporated with the rest of the demographic and personal data of the individual patient in a report.

POSTERS (PO)

Co-Chairs	Company
Meera Kumar	M & G Insights
Subrahmanyam (Gopal) Rajagopal	TAKE Solutions, Inc.

PO01 : LST in Comparison
Sanket Kale, Parexel
Sajin Johnny, Parexel
On Display – Meet the Poster Authors Mon. & Tues., 2:30 PM–3:30 PM, Chicago Ballroom Foyer

The need for producing error-free programming deliverables combined with the increasing globalization of the pharmaceutical industry has led to the growing requirements for validation of programs associated with generating clinical trial reports. Double programming is the most commonly accepted and widely used standard method in the industry for validation of programming algorithms and to cross-check adherence to specifications. Double programming involves two programmers programming independently based on specifications and then utilizing procedures like COMPARE to match the outputs. Currently the accepted QC process followed in most companies requires programmers to manually review the PROC COMPARE results for verifying the similarity of the outputs produced. This manual process can be very cumbersome and resource- intensive, especially while handling large deliveries, and also increases the likelihood of human error. This paper looks at an innovative method of eliminating this manual process with a SAS macro that scans the PROC COMARE output to check for similarity as well as listing any differences observed in the two outputs. The program will read in all .LST files in user-specified folders in a sequential manner and search the contents of each file for the existence of substrings to determine equivalence or list differences in the data compared. The derived results of all scanned .LST outputs will be summarized in one detailed report. The expectation is that this report will help improve the efficiency and quality of the validation process by eliminating the need for time-consuming manual review, thereby eliminating the risk of human error while also providing sufficient documentation in support of audit perspective processes.

PO02 : Derived Observations and Associated Variables in ADaM Data Sets
Arun Raj Vidhyadharan, inVentiv Health Clinical
On Display – Meet the Poster Authors Mon. & Tues., 2:30 PM–3:30 PM, Chicago Ballroom Foyer

We often produce ADaM datasets from their parent SDTM datasets. In this process, we see that the resultant ADaM datasets grow in size. It seems like an average-sized human being turned into a professional bodybuilder on steroids! So how do the ADaM datasets grow in size? Obviously not

steroids! When the datasets grow in width, we know that there are new variables that were not present in the parent SDTM datasets and we call them “derived variables”. And when the datasets grow in length, we know that there are new observations that came in and we call them “derived observations”. Some derived variables are associated with the derived observations and their purpose is solely to give more information on the derived observations. The focus of this paper is on the observations that we derive in ADaM data sets and the variables associated with those derived observations. The scope of this topic is well beyond the contents of this paper. There are limitless possibilities for deriving observations in ADaM datasets based on the requirements for reporting and other statistical analysis. The handful of scenarios described in this paper is based on my experience in my studies and the pharmaceutical companies that I worked with.

PO03 : A Drug Safety Reporting System in SAS
Yang Wang, Seattle Genetics
Shawn Hopkins, Seattle Genetics
Norm Fox, Seattle Genetics
Raghu Kumbharathi, Seattle Genetics
On Display – Meet the Poster Authors Mon. & Tues., 2:30 PM–3:30 PM, Chicago Ballroom Foyer

SAS is traditionally used as a primary programming language in clinical trial reporting, including tables, listings and figures (TLF) generation. Most of these TLFs are standard and designed to answer study questions on safety and efficacy of the drug during clinical trials, and for further submission and reporting to global regulatory agencies. This poster introduces a flexible and powerful drug safety reporting system that demonstrates the application of SAS in a non-traditional space. The drug safety database collects Serious Adverse Reactions (SAR) (including relatedness to study drug, dosing etc.) from clinical studies, investigator sponsored trials, post marketing surveillance and spontaneous reports. Reports created from this database are used to address multiple business needs. Drug Safety groups use these reports for signal detection, regulatory and safety reporting and in AE reconciliation with the clinical database. Medical Monitors use this for periodic AE monitoring and generating patient narratives. Clinical Operations groups use these reports for SAE Reconciliation with sites, and to monitor case status for IRBs. Additionally, various business reports are created to track case status, compliance and metrics. We implemented a process to streamline reporting by: 1) changing directory structure to accommodate unique requirements. 2) leveraging a common pool of analysis data sets to use across all reports. 3) building tools for automatic email notification to users when reports are ready. In the future, we plan to further automate the process to be able to create, validate and deliver a greater volume of reports more frequently and with fewer resources.

PO04 : Everything You Need To Know About Standardised MedDRA Queries
Rajkumar Sharma, Genentech, Inc.
On Display – Meet the Poster Authors Mon. & Tues., 2:30 PM–3:30 PM, Chicago Ballroom Foyer

MedDRA (Medical Dictionary for Regulatory Activities) is very commonly used in the pharmaceutical and biotech industry to classify adverse events information associated with the use of biopharmaceuticals and other products. In order to retrieve cases related to drug safety problems, users often combine MedDRA Preferred Terms (PT) from different System Organ Classes (SOCs). The challenge with this approach is that different users may select different sets of Preferred Terms to identify cases for the same drug safety problem. Also MedDRA has over 18,000 Preferred Terms, which makes it difficult to group events that indicate the presence of a condition that typically presents itself through different PTs. Standardized MedDRA Queries (SMQs) were developed to solve this problem. SMQs are groupings of MedDRA terms, ordinarily at the Preferred Term levels, that relate to a defined medical condition or area of interest. SMQs are intended to aid in the identification and retrieval of potentially relevant individual case safety reports. Since SMQs are driven by medical concepts and not by any specific company or product, they also eliminates any product-specific bias. This paper provides in-depth knowledge on Standardized MedDRA Queries and how to use them for identifying and retrieving safety data.

PO05 : ADaM Datasets for Graphs
Kevin Lee, Cytel
Chris Holland, Amgen
On Display – Meet the Poster Authors Mon. & Tues., 2:30 PM–3:30 PM, Chicago Ballroom Foyer

The paper is intended for clinical trial SAS® programmers who create graphs using ADaM (Analysis Data Model) datasets. ADaM datasets are analysis-ready, so programmers need to apply CDISC principles and guidelines to create them, while at the same time being mindful of their intent as plotting datasets. The paper will discuss the basic principles of graph creation using ADaM datasets. We propose the use of PARAM, AVAL, AVISIT and AVISITN for plotting. The paper will show in an example how to use AVAL for the y-axis and AVISITN for the x-axis for simple plotting. We will discuss the role of AVISITN as a numeric representation of AVISIT and propose the use of AVISITN for plotting purposes. We will also discuss some of the graphs created by features of SAS/STAT, ODS GRAPH and Graph Template Language (GTL). This will eliminate some of the intermediate SAS data sets that SAS programmers need to create for graph creation, which means that the graphs can be created directly from ADaM without the intermediate datasets. Examples using SAS/STAT, ODS GRAPH and GTL will be provided.

PO06 : Implementation of the Breast Cancer Risk Assessment Tool using SAS
Yuqin Li, Sr. SAS programmer
Lihua Chen, Statistics Manager
Xiaohai Wan, Senior Principal Statistical Scientist
Alan Chiang, Research Advisor
On Display – Meet the Poster Authors Mon. & Tues., 2:30 PM–3:30 PM, Chicago Ballroom Foyer

In this paper, a SAS macro is developed to implement the breast cancer risk assessment (BCRA) tool designed by National Cancer Institute (NCI). The BCRA tool itself is based on a complex statistical model known as the Gail model. The Gail model provides an estimate of a woman's risk of developing invasive breast cancer over a specific period of time by utilizing an individual's demographic information and risk factors. Breast cancer risk factors considered in the Gail model include (1) the number of previous breast biopsies, (2) the presence of atypical hyperplasia in any previous breast biopsy specimen, (3) age at the start of menstruation, (4) age at the first live birth of a child, (5) the history of breast cancer among her first-degree relatives (mother, sisters, daughters), and (6) the individual's age and race. The statistical model calculates individualized invasive breast cancer risk in terms of probabilities based on both the relative risk and the baseline hazard rate. We converted the C++ source code available from the NCI website to a SAS macro. Features of the macro include ease of implementation and integration through SAS, as well as flexibility in calculating the probabilistic breast cancer risk at any duration of time.

PO07 : Reliability Assessment of Image Data in Oncology and Psychology Studies
Li Zhang, Independent Consultant
David Shen, Independent Consultant
Gary Chen, Shire Pharmaceuticals
On Display – Meet the Poster Authors Mon. & Tues., 2:30 PM–3:30 PM, Chicago Ballroom Foyer

In oncology studies, the RECIST criteria provide the method for evaluating solid tumor response using MRI, CT scans and X-ray. By using brain-imaging-techniques like fMRI, cognitive psychology is able to analyze the relation between the physiology of the brain and mental processes. It is crucial to obtain accurate and reliable measurements in clinical studies. The analysts need to undergo comprehensive reliability assessment measures, including inter-reliability, prior to study commencement to verify measurement consistency and ensure a high level of agreement across analysts. The purpose of this study is to evaluate the proposed analysis methods to check for consistency and precision of measurements for imaging data.

PO09 : Efficient Statistical Review Using the ExcelXP Tagset
Bradford Danner, inVentiv Health Clinical
On Display – Meet the Poster Authors Mon. & Tues., 2:30 PM–3:30 PM, Chicago Ballroom Foyer

Working with clinical trial data, and assisting in the preparation of programming deliveries needed for clinical study reports, as statisticians we are often confronted with the review of many outputs. Part of such a review often requires spot checking of numbers. Programmers are frequently required to produce series of individual outputs, created in an iterative manner from one or few centralized programs, so that all are cosmetically consistent. Alternatively, as a statistician, our goal is to efficiently review several related outputs from one or few comparative outputs. We have found that the use of the REPORT and PRINT procedures presented using the ExcelXP tagset, and some of the built-in options, provides an effective tool to centralize and accelerate review of many outputs produced from a SAS program or SAS dataset.

PO10 : How to make SAS Drug Development More efficient
Xiaopeng Li, Celerion
Chun Feng, Celerion
Peng Chai, Celerion
On Display – Meet the Poster Authors Mon. & Tues., 2:30 PM–3:30 PM, Chicago Ballroom Foyer

SAS Drug Development (SDD) is a web-based SAS system on the UNIX operating system. Internet status, limited abilities to run multiple processes at a time and some SDD unique features take more time to operate a program in SDD. To run SDD efficiently is important for accomplishing tasks for programmers. In this paper, several methods that can make SDD more efficient will be examined.

PO11 : Flags for Facilitating Statistical Analysis Using CDISC Analysis Data Model
Chun Feng, Celerion
Xiaopeng Li, Celerion
Nancy Wang, Celerion
On Display – Meet the Poster Authors Mon. & Tues., 2:30 PM–3:30 PM, Chicago Ballroom Foyer

The CDISC Analysis Data Model (ADaM) is designed to be analysis-ready for statistical summary and analysis. Driven by analyses in the Statistical Analysis Plan (SAP), correct analysis flags should be set up in ADaM data sets prior to analyses. The ADaM subject-level analysis dataset (ADSL) contains subject-level population flags which indicate whether subjects are to be included in efficacy, safety, pharmacokinetic, pharmacodynamic, food effect, and dose proportionality analyses. In Basic Data Structure (BDS) data sets, common record-level analysis flags include, for example, re-check flags, flags for exclusion, baseline flags, early termination identifiers, and treatment emergent flags. However, from

our early phase clinical trial experience with ADaM, some analyses require additional customized analysis flags. For example, multiple baseline flags, average values as baselines flags, LOCF (last observation carried forward)/WOCF (worst value carried forward) flags, extra treatment breakdown flags for adverse events summarization, and analysis flags for sub-group analysis. In the paper, we first categorize basic key analysis flags and summarize their functions in a checklist. In addition, we will share some examples with customized analysis flags used in our practice. We will also extend the topic to what to consider while setting up analysis-ready ADaM data. The paper will prepare ADaM data users (implementers and reviewers) to check analysis flags appropriate for intended analyses.

PO12 : A SAS Users Guide to Regular Expressions When the Data Resides in Oracle
Kunal Agnihotri, PPD, Inc.
Kenneth Borowiak, PPD, Inc.
On Display – Meet the Poster Authors Mon. & Tues., 2:30 PM–3:30 PM, Chicago Ballroom Foyer

The popularity of the PRX functions and CALL routines has grown since they were introduced in SAS Version 9 due to the tremendous power they provide for matching patterns of text. Since the implementation of regular expressions within these functions is rooted in the Perl-style syntax, there is portability outside of a SAS environment. It is not uncommon for SAS users to access data residing in an Oracle-based environment. This paper explores the Oracle 10g implementation of regular expressions by highlighting similarities and differences to the PRX implementation in a series of queries using PROC SQL's pass-through facility against Oracle system tables.

PO13 : Traceability in the ADaM Standard
Ed Lombardi, SynteractHCR, Inc.
On Display – Meet the Poster Authors Mon. & Tues., 2:30 PM–3:30 PM, Chicago Ballroom Foyer

Traceability is one of the fundamentals of the ADaM Standard. However, there is not currently a traceability appendix to aid in implementation. Traceability is key to a truly ADaM-compliant submission. Having a clear path from SDTM to ADaM, within ADaM, and from ADaM to statistical analyses are all key components of ADaM datasets. This paper presents basic and advanced traceability techniques that can be useful for both reviewers and programmers. It also poses questions about difficult analyses for traceability such as occurrence and time-to-event analyses.

PO14 : V is for Venn Diagrams
Kriss Harris, SAS Specialists Ltd (Currently Contracting at Eli Lilly)
On Display – Meet the Poster Authors Mon. & Tues., 2:30 PM–3:30 PM, Chicago Ballroom Foyer

Would you like to produce Venn diagrams easily? This poster shows how you can produce stunning 2, 3 and 4-way Venn diagrams by utilizing the Graph Template Language, particularly the DRAWOVAL and DRAWTEXT statements. From my experience, Venn diagrams have typically been created in the pharmaceutical industry by using Microsoft Excel and PowerPoint. Excel is used to first count the numbers in each group and PowerPoint is used to generate the two- or three-way Venns. The four Way Venn diagram is largely unheard of, and when someone is brave enough to tackle it manually, then working out the numbers that should go in each of the 16 groups and inputting the right number into the right group is usually done nervously!

PO15 : SQL Subqueries: Usage in Clinical Programming
Pavan Vemuri, PPD
On Display – Meet the Poster Authors Mon. & Tues., 2:30 PM–3:30 PM, Chicago Ballroom Foyer

A feature of PROC SQL which provides flexibility to SAS users is that of a subquery. A subquery is a query inside of a query. A subquery can be further classified into an inline view, correlated subquery or uncorrelated subquery depending on the usage inside the query. This paper explores the types of subqueries, their merits and shortcomings, with examples related to clinical trials programming.

PO16 : TLFs: Replaying Rather than Appending
William Coar, Axio Research
On Display – Meet the Poster Authors Mon. & Tues., 2:30 PM–3:30 PM, Chicago Ballroom Foyer

In day-to-day operations of a Biostatistics and Statistical Programming department, we are often tasked with generating reports in the form of tables, listings, and figures (TLFs). Some requests come in the form of a small number of TLFs whereas others are more substantial in magnitude. Regardless, creating a single document for distribution and review may be required after all TLFs have been completed. A common setting in the pharmaceutical industry is to develop SAS code in which individual programs generate one or more TLFs in some standard formatted output such as RTF or PDF with a common look and feel. Furthermore, programs are developed over time, with the production run being in batch mode. The result is a set of TLFs completed at different times. Creation of a final (single) document with a properly sectioned and hyperlinked Table of Contents as well as dynamic page numbering may be desired. The ability to deliver a single document also greatly simplifies document management and electronic review for many end users. Many options have been proposed that post-process individual RTF or PDF files. An alternative approach which uses ODS Document will be introduced. Unlike many

other techniques, ODS Document uses intermediate files called template stores that are independent of ODS destination. Moderate success has been achieved with implementation of this technique in our specific setting. Details of this approach using SAS 9.2 with Windows will be discussed.

PO17 : Validating Listing Output: A Better Way
Hunter Vega, Stat-Tech Services, LLC
James Kniffen, Jr., Stat-Tech Services, LLC
On Display – Meet the Poster Authors Mon. & Tues., 2:30 PM–3:30 PM, Chicago Ballroom Foyer

When using the SAS System to create data listings for a clinical study, RTF output can range in size from one page to thousands of pages. In the past, this output was evaluated for accuracy using either the process of parallel programming and manual evaluation of selected records, or a comparison of output data sets using PROC COMPARE. While manual checks of long listings are demanding, they help identify errors such as truncation or incorrect formats that can be missed in a purely programmatic approach. The LISTCOMP macro extracts data directly from an RTF file, compares it to a parallel programmer's data set, and then provides a summary of each of the columns in the listings. Use of the macro is not dependent upon the version of SAS, but implementing the macro may require a moderate level of programming expertise. This approach not only ensures that the actual RTF output matches the parallel programmer's output, but that the reporting of information displayed in the RTF output allows for rapid review and identification of values that are impossible, incorrectly formatted, unformatted, or truncated. The LISTCOMP macro both reduces the validation time needed and increases the likelihood that problems in the output are identified.

PO19 : Navigating the Path to PDUFA V Conformance: 3 Strategies that are Critical to Success-Planningm Communicationand Governance
Joanna Koft, Biogen Idec, Inc.
On Display – Meet the Poster Authors Mon. & Tues., 2:30 PM–3:30 PM, Chicago Ballroom Foyer

With the recent approval of PDUFA V legislation, sponsors are both encouraged and hesitant of what lies ahead. For many, the day has come where upper management can finally be told that these data standards are required for submissions. For others, perhaps this is a blessing in disguise. For all, we now need to strategize and plan for our road to success in data standards implementation. This presentation will follow the 3 critical strategies for success: planning, communication, and governance. Having gone through multiple standards-based submissions, these three strategies have been identified as key to accomplishing the daunting task of conformance. I will discuss the three strategies including lessons learned and recommendations for future submissions.

PO20 : SAS Enterprise Guide® - Implementation Hints and Techniques for Insuring Success With Traditional SAS Programmers in a Pharmaceutical Development Role

Roger Muller, Data-to-Events.com

On Display – Meet the Poster Authors Mon. & Tues., 2:30 PM–3:30 PM, Chicago Ballroom Foyer

Traditional SAS programmers develop SAS code in files and submit it for processing regardless of the operating environment (PC, Unix, etc.). SAS Enterprise Guide (EG) follows this model, but adds some unique additional capabilities. This paper addresses set-up, initialization and workflow ideas to smoothen and enhance the transition to the EG graphical user interface centered around a process flow window. This paper will address local vs remote (server) processing. Starting with hardware and network capabilities, the paper then moves into a discussion on data location and how that affects workflow. Dual screen systems, split views and internal vs. external SAS code storage will be addressed. Certain features in SAS EG may either be hidden or exposed with advantages to doing either. The number of process flows in an Enterprise Guide Project offers flexibility in constructing the total programming effort. The ability to store code at a central facility as Stored Processes for sharing with other users will be discussed. Techniques for submitting developing code for step-by-step processing vs. the submission of entire project files is discussed. Backup strategies will be discussed. The paper will address environmental settings for both EG itself and for the advanced code editor. And lastly, the handiest key in EG, the F4 key, which is used to toggle back-and-forth between the current Process Flow window and the most recent window (program, data set, log, output, etc.) will be repeatedly emphasized. Configuration ideas for using the Query facility efficiently will also be presented.

STATISTICS AND PHARMACOKINETICS (SP)

Co-Chairs	Company
Venky Chakravarthy	BioPharma Data Services
Michael G. Wilson	Biostatistical Communications, Inc.

SP01 : Doctoring Your Clinical Trial with Adaptive Randomization
Jenna Colavincenzo, University of Pittsburgh
Monday, 9:15 AM - 9:35 AM, Location: Chicago Ballroom B

Adaptive randomization schemes have become increasingly common in beginning stages of clinical trials and in small clinical trials. This paper introduces two kinds of adaptive randomization schemes - treatment adaptive randomization and covariate adaptive randomization - and discusses the benefits and limitations of each. In addition, this paper demonstrates how to use SAS macros (written with SAS 9.2) to perform these adaptive randomization schemes in a clinical setting, and how these macros can be modified to fit your randomization needs.

SP02 : Probability Based Criteria for Early Phase Drug Development
Howard Liang, inVentiv Health Clinical
Monday, 9:45 AM - 10:05 AM, Location: Chicago Ballroom B

Probability based criteria lies in the study design where many research scientists prefer to understand the power and probability of taking a new drug forward across the whole range of possible true treatment effects, rather than focusing on one particular value to power the study. Examples will be used in this paper to show how to compute probability using the SAS/STAT procedure PROC MIXED. Particular emphasis is given to its application on efficacy analysis, including a comparison to classical hypothesis testing. The application on safety analysis is also discussed. SAS products used include SAS/STAT: PROC MIXED; BASE SAS: PROC REPORT and PROC SQL. Operating system and SAS version: no dependency. Audience skill level: basic statistical concepta and some experience using PROC REPORT, PROC SQL AND PROC MIXED

SP03 : Combining Analysis Results from Multiply Imputed Categorical Data
Bohdana Ratitch, Quintiles
Ilya Lipkovich, Quintiles
Michael O'Kelly, Quintiles
Monday, 10:15 AM - 11:05 AM, Location: Chicago Ballroom B

Multiple imputation (MI) is a methodology for dealing with missing data that has been gaining wide usage in clinical trials. Various methods are readily available in SAS PROC MI for multiple imputation of both continuous and categorical variables. MI produces multiple copies of the original data set, where missing data are filled in with values that differ slightly between imputed data sets. Each of these data sets is then analyzed using a standard statistical method for complete data, and the results from all imputed data sets are combined for overall inference using Rubin's rules which account for the uncertainty associated with imputed values. Rubin's pooling methodology is very general and is essentially the same no matter what kind of statistic is estimated at the analysis stage for each imputed data set. However, the combination rules assume that the estimates are asymptotically normally distributed, which may not always be the case. For example, the Cochran-Mantel-Haenszel (CMH) test and the Mantel-Haenszel (MH) estimate of the common odds ratio are often used in analysis of categorical data, and they produce statistics that are not normally distributed. In this case, normalizing transformations need to be applied to the statistics estimated from each imputed data set before the Rubin's combination rules can be applied. In this paper, we show how this can be done for the two aforementioned statistics and explore some operating characteristics of the significance tests based on the applied normalizing transformations. We also show how to obtain combined estimates of binomial proportions and their difference between treatment arms.

SP04 : Adjusted Proportion Difference and Confidence Interval in Stratified Randomized Trials
Yeonhee Kim, Gilead Sciences
Seunghyun Won, University of Pittsburgh
Monday, 1:15 PM - 1:35 PM, Location: Chicago Ballroom B

Stratified randomization is widely used in clinical trials to achieve treatment balance across strata. In the analysis, investigators are often interested in the estimation of common treatment effect adjusting for stratification factors. We developed the macro for stratified analysis when the primary endpoint is a difference in proportions between two treatment groups. The %strataCI macro estimates proportion difference using three different weighting schemes and calculates both Wald and Newcombe confidence intervals.

SP05 : Case-Control: The Analysis of Biomarker Data Using SAS Genetic Procedure
Jaya Baviskar, InVentiv Health Clinical
Monday, 2:15 PM - 2:35 PM, Location: Chicago Ballroom B

Genetics aids to identify any susceptible diseases at the root cause level due its inherent ability to provide an array of information very specific to the building blocks of life. Hence, analysis of genetic data like biomarkers or the genetic make-up is a crucial task and to derive accurate inferences adds complexity to it further. Taking this into consideration, SAS has introduced new procedures that help analyze genetic data so as to arrive at accurate conclusions. The CASECONTROL procedure is designed to handle such biomarker data, thereby helping to analyze and assess data more effectively and efficiently. Backed with statistical concepts and built-in options, the procedure allows us to focus more on the interpretation of data in question through 3 readily available Chi-squared tests. The paper will discuss bringing data into the desired format, applying SAS options available, statistical computations to be considered, and deriving correct inferences.

SP07 : Time to Event Analysis in the Pharmaceutical and Medical Device Industries
Helen Chmiel, Experis
Evan Ritzema, Experis
Monday, 1:45 PM - 2:05 PM, Location: Chicago Ballroom B

Time to event analysis, often referred to as survival analysis, is standard in many industries to evaluate clinically meaningful time to event outcomes, such as time to implant failure or death. In addition to this analysis being widespread, specific needs are varied based on the industry and require software with the ability to generate the required results as well being flexible and easy to use. Some examples of the current industry need for time to event analysis includes survival analysis for human subjects in a clinical trial environment, time to failure analysis in both cardiac stent and orthopedic device implantation, and time to death toxicology profiles for a dose level study in preclinical studies. With this great need for reliable measures to make risk/benefit assessments in the pharmaceutical and medical device industries, there is a corresponding great need for software applications for analysts and clinicians. This presentation is an application which illustrates how SAS macro processing can be used with the statistical procedures of LIFETEST and PHREG to provide the flexibility needed for time to event analysis. This application can further be used to illustrate how SAS DATA step processing can be used for data preparation and graphics production in two different scenarios of time to event analyses. This presentation will be most valuable for experienced SAS programmers with exposure to macro processing.

SP08 : Oh Quartile, Where Art Thou?
David Franklin, TheProgrammersCabin.com
Monday, 11:15 AM - 11:35 AM, Location: Chicago Ballroom B

"Why is my first quartile number different from yours?" It was this question that led to a study of methods that can be used for calculating the first and third quartile for a set of data. SAS provides five methods for calculation of the quartile but other methods exist and are used in other common software packages. In this paper ten methods are discussed from the point of view of a SAS programmer and a macro presented that will calculate first and third quartiles for the five methods not provided in the Base SAS package.

SP09 : SAS 9.3: Better Graphs, Easier Lives for SAS Programmers, PK Scientists and Ppharmacometricians
Alice Zong, M.S.
Monday, 2:45 PM - 3:05 PM, Location: Chicago Ballroom B

Data visualization tools are widely used in PK analysis and PK/PD modeling and simulation. To date, even though the data sets are mostly in SAS data format, PK scientists and pharmacometricians usually choose to use other software for the production of graphs; however, this approach often requires changing data types and repeatedly transferring data between functional teams. With the new release of SAS 9.3 and the enhancement of SAS graphs, it is now much easier for SAS programmers to make graphs to support analysis, modeling and simulation activities. This paper illustrates some commonly used graphs that are currently generated in other software applications, e.g. R/S-PLUS, that can now be generated in SAS. This change has the potential to significantly improve the process and productivity of model-support/based drug development.

SP10-SAS : Introduction to Bayesian Analysis Using SAS® Software
Maura Stokes, SAS
Tuesday, 8:00 AM - 9:50 AM, Location: Los Angeles/Miami

Bayesian methods have become increasingly popular in recent years in a number of different disciplines. This tutorial provides an introduction to Bayesian methods with applications in the areas of the generalized linear model and survival analysis. The first part of the course provides an overview of Bayesian methodology, including motivation and Bayesian inference, as well as computational methods and convergence diagnostics relevant to the SAS implementation. The second part of the course discusses applications using new capabilities in SAS/STAT software in the GENMOD, LIFEREG and PHREG procedures which are based on Gibbs sampling. Examples will include methods such as linear regression, logistic regression, Poisson regression, Cox regression, parametric survival models, and the piecewise exponential model. A master's level knowledge of statistics is assumed as well as experience with generalized linear models and survival analysis. Previous exposure to Bayesian methods is useful but not required.

TECHNIQUES & TUTORIALS: FOUNDATIONS (TF)

Co-Chairs	Company
Rob Howard	Veridical Solutions
Kirk Paul Lafler	Software Intelligence Corp.

TF01 : Why the Bell Tolls 108 Times? Stepping Through Time with SAS
Peter Eberhardt, Fernwood Consulting Group Inc.
Tuesday, 2:15 PM - 2:35 PM, Location: Chicago Ballroom G

For many SAS programmers, new or even advanced, the use of SAS date and datetime variables is often very confusing. This paper addresses the problems that the most of programmers have. It starts by looking at the basic underlying difference between the data representation and the visual representation of date, datetime and time variables. From there it discusses how to change data representations into visual representations through the use of SAS formats. Since date manipulation is core to many business process, the paper also discusses date arithmetic first by demonstrating the use of simple arithmetic to increment dates; then by moving on to SAS functions which create, extract and manipulate SAS date and datetime variables. Finally, the paper demonstrates the use of the %SYSFUNC macro function and the %LET statement to present date, datetime and time variables. This paper is introductory and focuses on new SAS programmers;, however, some advanced topics are also covered.

TF02 : Things Dr Johnson Did Not Tell Me: An Introduction to SAS® Dictionary Tables
Peter Eberhardt, Fernwood Consulting Group Inc.
Wednesday, 9:30 AM - 9:50 AM, Location: Chicago Ballroom G

SAS maintains a wealth of information about the active SAS session, including information on libraries, tables, files and system options; this information is contained in the dictionary tables. Understanding and using these tables will help you build interactive and dynamic applications. Unfortunately, dictionary tables are often considered an advanced topic by SAS programmers. This paper will help novice and intermediate SAS programmers get started with their mastery of the dictionary tables. Ever needed a list of the tables (datasets) in a library? How about the columns (variables) in a table? Need to make sure you reset any titles after you run a report? Got some pesky warning messages in your SAS log you would like to clean up? Sure, you can look them up in the table and column properties in the Explorer window. Or you can run a PROC CONTENTS and check the listing. And of course you can ignore the warnings and errors in the SAS log since they almost always appear. Or you can go to the dictionary tables and have your programme find out what libraries are allocated or what columns are available. So, what are dictionary tables and where do we access them?

TF03 : The SAS DATA Step: Where Your Input Matters
Peter Eberhardt, Fernwood Consulting Group Inc.
Monday, 4:00 PM - 4:50 PM, Location: Chicago Ballroom G

Before the warehouse is stocked, before the stats are computed and the reports run, before all the fun things we do with SAS can be done, the data need to be read into SAS. A simple statement, INPUT, and its close cousins FILENAME and INFILE, do a lot. This paper will show you how to define your input file and how to read through it, whether you have a simple flat file or a more complex formatted file.

TF04 : There's Nothing ODiouS About ODS
Aaron Rabushka, inVentiv Health Clinical
Wednesday, 9:00 AM - 9:20 AM, Location: Chicago Ballroom G

This presentation provides a look at SAS ODS capabilities with regard to enhancing output, suppressing output, and drawing information from SAS PROCs into working datasets. Its code examples were tested under SAS V9.1.3, SP4.

TF05 : Making the Log a Forethought Rather Than an Afterthought
Emmy Pahmer, inVentiv Health Clinical
Tuesday, 4:00 PM - 4:20 PM, Location: Chicago Ballroom G

When we start programming, we simply hope that the log comes out with no errors or warnings, yet once we have programmed for awhile, especially in the area of pharmaceutical research, we realize that having a log with specific useful information in it improves quality and accountability. We will discuss clearing the log, sending the log to an output file, helpful information to put in the log, which messages are permissible, automated log checking, adding messages regarding data changes, whether or not we want to see source code, and a few other log-related ideas. Hopefully, the log will become something that we keep in mind while we're programming, rather than just an afterthought.

TF06 : Let SAS® Do Your DIRty Work
Richann Watson, Experis
Monday, 2:45 PM - 3:05 PM, Location: Chicago Ballroom G

Making sure you have all the necessary information to replicate a deliverable saved can be a cumbersome task. You want to make sure that all the raw data sets are saved, all the derived data sets, whether they are SDTM or ADaM data sets, are saved and you prefer that the date/time stamps are preserved. Not only do you need the data sets, you also need to keep a copy of all programs that were used to produce the deliverable as well as the corresponding logs when the programs were executed.

Any other information that was needed to produce the necessary outputs needs to be saved. All of this needs to be done for each deliverable and it can be easy to overlook a step or some key information. Most people do this process manually and it can be a time-consuming process, so why not let SAS do the work for you?

TF08 : Google® Search Tips and Techniques for SAS® and JMP® Users
Charles Edwin Shipp, Consider Consulting Inc.
Kirk Paul Lafler, Software Intelligence Corporation
Monday, 3:30 PM - 3:50 PM, Location: Chicago Ballroom G

Google (www.google.com) is the world's most popular and widely-used search engine. As the premier search tool on the Internet today, SAS and JMP users frequently need to identify and locate SAS and JMP content wherever and in whatever form it resides. This paper provides insights into how Google works and illustrates numerous search tips and techniques for finding articles of interest, reference works, information tools, directories, PDFs, images, current news stories, user groups, and more to get search results quickly and easily.

TF09 : Give the Power of SAS to Excel Users Without Making Them Write SAS Code
William E Benjamin Jr, Owl Computer Consultancy LLC
Tuesday, 9:00 AM - 9:50 AM, Location: Chicago Ballroom G

Merging the ability to use SAS and Excel can be challenging. However, with the advent of SAS Enterprise Guide, SAS Integration Technologies, SAS BI Server software, SAS JMP software, and SAS Add-ins for Microsoft Office products; this process is less cumbersome. Using Excel has the advantages of being cheep, available, easy to learn, and flexible. On the surface SAS and Excel seem widely separated without these additional SAS products. But wait, BOTH SAS AND EXCEL CAN INTERFACE WITH THE OPERATING SYSTEM. SAS can run Excel using the X command and Excel can run SAS as an APPLICATION. This is NOT DDE; each system works independent of the other. This paper gives an example of Excel controlling a SAS process and returning data to Excel.

TF10 : Extend the Power of SAS to Use Callable VBS and VBA Code Files Stored in External Libraries to Control Excel Formatting Routines
William E Benjamin Jr, Owl Computer Consultancy LLC
Wednesday, 10:45 AM - 11:05 AM, Location: Chicago Ballroom G

Did you ever wish you could use the power of SAS to take control of Excel and make Excel do what you wanted WHEN YOU WANTED? Well, one letter is the key to doing just that, the letter X as in the SAS X Command that opens the door to all operating system commands from SAS. The Windows operating system comes with a facility to write a series of commands called scripts. These scripts have the ability

to open and reach into the internals of Excel. Scripts can load, execute and remove VBA macro code and control Excel. This level of control allows you to make Excel do what you want, without leaving any traces of a macro behind. This is Power.

TF11 : "How Do I . . .?" There is More Than One Way to Solve that Problem; Why Continuing to Learn is So Important
Arthur Carpenter, CALOXY
Tuesday, 1:15 PM - 2:05 PM, Location: Chicago Ballroom G

In the SAS forums questions are often posted that start with "How do I . . . ?". Generally there are multiple solutions to the posted problem, and these vary from simple to complex. All too often the simple solution is both inefficient and reflects a naïve understanding of the SAS language. This would not be so very bad except sometimes the responder thinks that their response is the best solution or perhaps worst the only solution. Worse yet, when there is a range of solutions, the 'right answer' that the original poster selects often reflects the simplest solution that the original poster understands. In both cases these folks have stopped learning and have stopped expanding their understanding of the language. The examples in this presentation will illustrate the progression of solutions from the simple (simplistic) to the sophisticated for a number of "How do I . . . ?" questions, and through the discussion of the individual techniques, we will learn how and why it is so very important to continue to learn.

TF12 : Let's get SASsy
Amie Bissonett, inVentiv Health Clinical
Wednesday, 8:00 AM - 8:20 AM, Location: Chicago Ballroom G

The SAS language has a plethora of procedures, DATA step statements, functions, and options to boot. Most programmers with years of experience have yet to use all that SAS has to offer. This paper gives a variety of tips and introduces a few handy tools to add to your SAS arsenal, including good programming practices, uses of the RETAIN statement, and macro tips.

TF13 : Mission Possible: Your Assignment, Should You Decide to Accept It, Is to Validate a Study
Susan Fehrer Coulson, BioClin, Inc.
Tuesday, 2:45 PM - 3:05 PM, Location: Chicago Ballroom G

You have been called into your manager's office and told that your next assignment is to start validating the programming in a study, that not necessarily has been one that you have worked on. Other than checking your own programs and output, where do you start? And, why does this task have to be done? It can be a daunting task to start validating a project, any project, if it is the one you worked on or not. And, where is the best place to start? With examples and best practices on where and how to start, you will not have a Mission Impossible.

TF15 : What's Hot, What's Not - Skills for SAS Professionals
Kirk Paul Lafler, Software Intelligence Corporation
Charles Edwin Shipp, Consider Consulting
Tuesday, 5:00 PM - 5:20 PM, Location: Chicago Ballroom G

As a new generation of SAS user emerges, current and prior generations of users have an extensive array of procedures, programming tools, approaches and techniques to choose from. This presentation identifies and explores the areas that are hot and not-so-hot in the world of the professional SAS user. Topics include Enterprise Guide, PROC SQL, JMP, PROC REPORT, Macro Language, ODS, DATA step programming techniques such as arrays and hashing, sasCommunity.org, LexJansen.com, JMP, and Output Delivery System (ODS).

TF17 : Essentials of PDV: Directing the Aim to Understanding the DATA Step!
Arthur Li, City of Hope
Monday, 1:15 PM - 2:05 PM, Location: Chicago Ballroom G

Beginning programmers often tend to focus on learning syntax without understanding how SAS processes data during the compilation and execution phases. SAS creates a new data set, one observation at a time, from the program data vector (PDV). Understanding how and why each of the automatic or user-defined variables is initialized and retained in the PDV is essential for writing an accurate program. Among these variables, the following variables deserve special attention, including variables that are created in the DATA step, by using the RETAIN or the SUM statement, and via BY-group processing (FIRST.VARIABLE and LAST.VARIABLE). In this paper, you will be exposed to what happens in the PDV and how these variables are retained from various applications.

TF18 : The By-laws of BY-Group Processing
Tracee Sorrentino, Consulting Analyst
Tuesday, 8:00 AM - 8:20 AM, Location: Chicago Ballroom G

One of the big challenges with health care claims data is that they contain multiple rows of transactions, e.g. rows, per patient, per visit, depending on the services rendered that day. Usually patients and their diseases are studied over time, and for this reason BY-group processing is a mandatory tool. This paper will showcase the power of BY-group processing and will also cover what happens within SAS, behind the scenes. The process will be illustrated with an example of its usefulness by creating a physician targeting list, with metrics, where physicians practice in more than one location.

TF19 : SQL, HASH Tables, FORMT and KEY= - More Than One Way to Merge Two Data Sets
David Franklin, TheProgrammersCabin.com
Tuesday, 3:30 PM - 3:50 PM, Location: Chicago Ballroom G

The MERGE statement is the most common way to merge one-to-one or one-to-many data. This works very well most of the time but there are other methods that are useful, and sometime more efficient, that should be every SAS programmer's toolbox. This paper touches on four methods that can be more efficient: a quick look at PROC SQL and some of the options that help, HASH tables and some of the considerations for using this format, PROC FORMAT, and the KEY= option in the SET statement.

TF20 : Working Hard to Become Lazy
Sunil Kumar Pusarla, Omeros Corporation
Monday, 5:00 PM - 5:20 PM, Location: Chicago Ballroom G

SAS programmers in pharmaceutical/biotech industry usually write similar types of code in different study programs. Re-usability of the code we produce makes our life easier. Many situations arise when we need to modify our own code. In such times, rather than inspecting each line of a program for where to modify (this is a huge amount of work and we may easily miss something), it is better when our modifications are constrained to one easily identified location, which affects all subsequent code automatically. There are many simple yet effective ways this will increase our productivity and speed. In this paper, I would like to discuss some of those techniques (including legacy procedures like FORMAT, new enhancements to existing procedures like PROC SUMMARY, obscure functions, as well as more recent resources such as PROC FCMP).

TF21 : Access to Relational Databases Using SAS
Frederick Pratter, Utopia Ltd.
Wednesday, 10:15 AM - 10:35 AM, Location: Chicago Ballroom G

SAS software currently provides many of the features of a database management system, including database views and an extended superset of ANSI SQL. However, it is often impractical or just plain impossible to convert desktop or organizational databases into SAS. Consequently, SAS software provides several procedures for access to relational databases. This paper will review how to use the various SAS/ACCESS products for linking networked workstations to remote servers. Most of the examples will use the SAS/ACCESS for Oracle, but the principles described apply equally to local databases in Microsoft Access, as well as other client/server systems such as Netezza, DB2 and MySQL.

TF22 : Anatomy of a Merge Gone Wrong
James Lew, Compu-Stat Consulting
Joshua Horstman, Nested Loop Consulting
Tuesday, 8:30 AM - 8:50 AM, Location: Chicago Ballroom G

The merge is one of the SAS programmer's most commonly used tools. However, it can be fraught with pitfalls to the unwary user. In this paper, we look under the hood of the DATAstep and examine how the Program Data Vector works. We see what's really happening when data sets are merged and how to avoid subtle problems.

TF23 : Effective Independent Validation - Tips to Improve the Independent Validation Process
Daniel Butner, PPD
Brandon Graham, PPD
Wednesday, 8:30 AM - 8:50 AM, Location: Chicago Ballroom G

Independent validation is the act of two programmers working independently from a common set of specifications and comparing their results. This process enhances quality by ensuring two qualified programmers come to the same analysis result before it is considered "validated". Independent validation is commonly used in the pharmaceutical/CRO industry. Skilled SAS programmers who are effective at independent validation can greatly increase the quality and efficiency of a project. This paper will examine the technical approaches that may be used to identify differences between independent programmers. It will include approaches to validating analysis data sets as well as standard output. In addition, specific techniques to improve efficiency and detail will be discussed. A discussion of providing validation feedback will also be included. It will discuss considerations with regard to the amount of detail provided and methods for providing feedback, and offer tips for communicating effectively during the validation process.

TF24 : Defensive Programming and Error-handling: The Path Less Travelled
Tracy Sherman, InVentiv Health Clinical
Angela Ringelberg, InVentiv Health Clinical
Tuesday, 4:30 PM - 4:50 PM, Location: Chicago Ballroom G

Defensive programming detects logic errors by determining whether the code represents a proper implementation of the specifications. Thus, defensive programming is a prospective approach that anticipates logic errors, even those not yet discovered. Once a logic error is identified, messages are written to the SAS log and, possibly, additional diagnostics commences. The process emulates a thought experiment that reduces logic errors. This paper explains defensive programming and error-handling techniques based on the implementation of CDISC ADaM (analysis) data sets, which can be easily extended to other programming tasks.

TF25-SAS : Creating and Customizing the Kaplan-Meier Survival Plot in PROC LIFETEST

Warren Kuhfeld, SAS

Ying So, SAS

Monday, 2:15 PM - 2:35 PM, Location: Chicago Ballroom G

If you are a medical, pharmaceutical, or life sciences researcher, you have probably analyzed time-to-event data (survival data). One of several survival analysis procedures that SAS/STAT® provides, the LIFETEST procedure computes Kaplan-Meier estimates of the survivor functions and compares survival curves between groups of patients. You can use the Kaplan-Meier plot to display the number of subjects at risk, confidence limits, equal-precision band, Hall-Wellner band, and homogeneity test p-value. You can control the contents of the survival plot by specifying procedure options with PROC LIFETEST. When the procedure options are insufficient, you can modify the graph template with SAS macros. This paper provides basic examples of survival plots, survival plot modification using the procedure options, graph template modifications using the macros, and more advanced examples.