

# BASUG Quarterly Meeting Announcement

## Short Meeting Description

The non-profit Clinical Data Interchange Standards Consortium (aka CDISC) continues to develop data standards that are being warmly embraced by the FDA. Come learn about streamlining the implementation of these standards. Resistance is futile.

This meeting is particularly relevant to SAS programmers, Biostatisticians, Clinical Data Managers, and Regulatory Specialists in the Biotechnology/Pharmaceutical and Medical Device industries. Those outside clinical research in these industries may find this meeting helpful for informing development of standards for their own data.

*Immediately following the meeting, we will provide an informal light buffet (free) lunch for all meeting attendees. We hope you can stay for this opportunity to network and socialize with your fellow SAS users.*

## Additional Background

The mission of the Clinical Data Interchange Standards Consortium (CDISC) is to “develop and support global, platform-independent data standards that enable information system interoperability to improve medical research and related areas of healthcare.” Since its formation as an independent, non-profit in 2000, CDISC has made substantial progress towards its mission and has developed a suite of data and metadata standards that are being adopted by the Food and Drug Administration (FDA).

The ultimate goal of data standardization across organizations is to maximize efficiency and quality of the FDA submission process overall. Often, however, efficiency is initially reduced until processes are developed and streamlined to fully implement standards. The speakers will discuss challenges and opportunities related to implementing two CDISC standards, the Study Data Tabulation Model (SDTM) and Case Report Tabulation Data Definition Specification using define.xml.

Given that the FDA has adopted SDTM and define.xml for submitting tabulation data it is inevitable that all submissions will eventually be expected to conform to these standards. Thus it is imperative for those working with clinical data in the Biotechnology and Pharmaceutical industry to become proficient in these standards.

**Topic:** Managing Clinical Data in the Age of CDISC

**When:** March 2, 2011  
8:15am – Noon

**Where:** Microsoft New England Research & Development Center  
One Memorial Drive  
Conference Center First Floor  
Cambridge, MA 02142  
(857) 453-6000

**Directions:** Please visit the [meeting site directions page](#).

**How:** **Individual, On-Line Registration Required. No Email!**

**To register:** Please visit the [meeting registration page](#).

**Contact:** If you have questions about the meeting, contact [Asli Memisoglu](#) or [Manjusha Gokhale](#)

## Agenda\*

8:15 Sign in and Refreshments

9:00 Announcements

9:15 In-Depth Review of Validation Tools to Check Compliance of CDISC SDTM-Ready Clinical Datasets by Bhavin Busa, Sr. Statistical Programmer, Cubist Pharmaceuticals, Inc.

10:30 Break

10:45 The Use of Metadata in Creating, Transforming and Transporting Clinical Data  
by Gregory Steffens, Director, Data Management and SAS Programming, ICON Development Solutions

Noon Meeting Adjourned

Noon-1:00pm FREE Networking Lunch

*\*Note: Times (and sequence) are approximate and subject to change. Please re-visit [the BASUG website](#) for updated information.*

## Abstracts and Bios

**In-Depth Review of Validation Tools to Check Compliance of CDISC SDTM-Ready Clinical Datasets** by Bhavin Busa, Sr. Statistical Programmer, Cubist Pharmaceuticals, Inc.

### Abstract

As the pharmaceutical organizations becomes increasingly involved in developing efficient and cost effective ways to produce CDISC SDTM-compliant clinical trial domains, it has become more crucial to identify a validation tool to check compliance and streamline operations for the preparation of submission-ready files in accordance with the most recent SDTM Implementation Guide. The ultimate need of a SDTM validation tool is to check the compliance of the SDTM domains for successful load into Janus (clinical data repository at FDA) and thereby reducing risk of delay in the submission review process.

The presentation will provide an in-depth review of various SDTM validation tools and provide some insight on the framework, installation and pros & cons of each solution by implementing them on a real submission-ready clinical datasets. The high-level outline of the presentation is as follows:

- Brief introduction of CDISC SDTM standards
- FDA initiatives on CDISC SDTM standards
- SDTM Data Flow from Sponsor to FDA
- Needs and requirements of an SDTM validation tool

- In-depth Review of various SDTM validation tools:
  - o Lincoln Technologies WebSDM™
  - o SAS® – Clinical Standard Toolkit
  - o OpenCDISC Validator
  - o In-House SAS Macro Based Solution
- Side-by-Side comparison of SDTM validation tools
- Summarizing evaluation results and recommendation
- Open the floor for discussion on other implementation ideas and strategies

### Speaker Bio

Bhavin Busa is a Senior Statistical Programmer at Cubist Pharmaceuticals. He is an active participant in CDISC initiative and is well adept with SDTM, ADaM, and Define.xml standards. He is a regular presenter at Pharmaceutical SAS User Group, Mass Biotech Council, and is a committee member of Boston Area CDISC User Network. In the past, he has been honored with the best paper award in the Regulatory Submission section of the PharmaSUG 2008 on the very topic of CDISC SDTM validation. In addition to his expertise in CDISC SDTM, he has also developed various SAS based solution to automate creation of analysis tables and Define.xml. On a personal note, he believes in sharing knowledge and is therefore very active in presenting topics that will benefit the SAS community and the industry as a whole.

### **The Use of Metadata in Creating, Transforming and Transporting Clinical Data**

by Gregory Steffens, Director, Global SAS Programming, ICON Development Solutions

### Abstract

Pharmaceutical companies must address the need to define database structures quickly and easily in order to

- support clinical trial analysis databases
- import data from central laboratories and CROs
- export data to Data Safety Monitoring boards
- share data between corporate sites within and without the United States
- submit data to the FDA and other regulatory agencies

Industry-level standards are being discussed and developed to help address this need and most pharmaceutical companies have corporate standards of database structures. However, generally these industry and corporate standards are stored in .pdf or MSword documents. As such they are accessible only to people reading the documents. Storing this same information in meta data sets adds much value to the effort put into defining these standards because this information is now available to computer programs. This presentation describes standard meta data set structures that are capable of storing specifications for any of the above database requirements. These meta data, along with SAS macros and a SAS/AF application that access these meta data, supports the specification, creation, importation, exportation, comparison and validation of databases and format catalogs. Further, these meta data can be used to export data and metadata to XML format. The meta data structures can be implemented in any relational database or in SAS itself. The macros support such functions as:

- printing the specification in several formats, including the define.xml and define.html formats
- adding all the data set and variable attributes to the database
- listing discrepancies between the specification and the database
- sorting the data by primary keys
- reordering variables according to FDA preferences

- creation of format catalogs
- creation of character decode variables from code variables
- comparison of database structures to assist in enforcing project standards
- transforming a database from one database structure to another, using map metadata

As industry and corporate standards are developed there is great value in documenting these in meta data sets, that are accessible to computer programs, rather than in word or .pdf.

In summary, when importing data from outside sources, the data specification is implemented in meta data so that computer programs can automatically compare the data submitted by the lab or CRO to the specification and list discrepancies where the database does not conform to the specification. When specifying study analysis databases, macros assist in the building of the database by automatically adding all the specified variable and data set attributes, such as labels and formats, adding decode variables, sorting the data, building format catalogs, etc. When submitting data to the FDA a thorough specification can be printed from the metadata or exported to XML format, with bookmarks and hyperlinks.

#### Speaker Bio

Greg Steffens has been using SAS for programming and applications development since 1981, primarily in the pharmaceutical and health insurance industries. He has held job positions ranging from lead technical to director-level management. He is currently Director of SAS Programming at ICON PLC and a member of two CDISC teams. Greg's experience includes the design and development of metadata and software to automate data definition, data transformation, data validation and FDA submissions.

## **BASUG Membership**

Keep your BASUG Membership up-to-date! [For more information on our membership policy or to purchase a membership, visit the BASUG membership page](#)

## **Directions to Microsoft New England Research & Development Center**

Please visit the [meeting site directions page](#).

## **BASUG Contacts**

Mailing Address:

BASUG  
PO Box 170253  
Boston MA 02117

Email our [Webmaster](#)

# BASUG Quarterly Meeting Announcement

Move yourself to a higher level of programming by making use of “metadata.” What is “metadata?” It is data *about* data. By understanding and using metadata, your applications will be much more flexible and extensible. See below for details.

Please join us for these informative talks, and consider staying for the afternoon class on metadata (separate event). For more information on the afternoon training, please visit [the training announcement](#).

*Immediately following the meeting, we will provide an informal light buffet lunch for all meeting attendees. We hope you can stay for this opportunity to network and socialize with your fellow SAS users.*

**Topic:** Metadata in SAS Applications

**When:** **Wednesday, June 15, 2011**  
**8:15am – Noon**

**Where:** **Microsoft New England Research & Development Center**  
One Memorial Drive  
Conference Center First Floor  
Cambridge, MA 02142  
857-453-6000

**Directions:** Please visit the [meeting site directions page](#).

**How:** **Individual, On-Line Registration Required. No Email!**

**To register:** Please visit the [event registration page](#).

**Contact:** If you have questions about the meeting contact:  
[Bruce Haimowitz](#)  
[Robert Rosofsky](#)

## Agenda\*

- 8:15 Sign in and Refreshments
- 8:45 Announcements
- 9:00 Metadata: Some Fundamental Truths  
by Frank Dilorio, CodeCrafters, Inc.
- 9:50 Break
- 10:05 Using Datasets to Define Macro Loops and Local Macro Variables  
by Steven Ruegsegger, IBM Microelectronics, Burlington, VT
- 11:00 Break
- 11:10 Building the Better Macro: Best Practices for the Design of Reliable,  
Effective Tools  
by Frank Dilorio, CodeCrafters, Inc.
- Noon Meeting Adjourned – Networking Lunch

*\*Note: Times (and sequence) are approximate and subject to change. Please re-visit [the BASUG website](#) for updated information.*

## Abstracts and Bios

### **Metadata: Some Fundamental Truths**

by Frank Dilorio

One characteristic of a healthy workplace is its ability to take on increased and more diverse work loads without compromising product quality. Workflow volume and quality can often be addressed by adding to existing staff and improving the computing environment. At some point, however, usually when the work flow becomes a work deluge, the need for a paradigm shift becomes apparent. Metadata – data describing corporate processes and data – is a potent framework for making this shift.

This paper discusses a number of conceptual issues related to the design, implementation, and growth of metadata-based systems. It identifies situations where metadata can improve processes and suggests how to evaluate both the benefits and costs of implementation. The treatment of the topic is high-level. The reader will not learn coding and design techniques, but will gain an appreciation of the power of metadata-driven workflow, and will have an eyes wide open understanding of what resources need to be expended to achieve it. Although the scenarios used are from the pharmaceutical industry, the larger, take-away points are applicable to all sectors.

Frank Dilorio is President of CodeCrafters, Inc., a consulting firm specializing in pharmaceutical applications and SAS training. A SAS programmer since 1975, he is the author of *SAS Applications Programming: A Gentle Introduction* and *Quick Start to Data Analysis with SAS*. He is a frequent speaker at local, regional, and international SAS conferences, presenting at last count over 200 papers and workshops.

Frank is past President of the SouthEast SAS Users Group, and was co-chair of its 1994 and 1996 conferences. He continues to be active in the regional and local SAS user groups and was a co- founder of the Research Triangle CDISC Users Group.

### **Using Datasets to Define Macro Loops and Local Macro Variables**

by Steve Ruegsegger

A common and optimal way to execute repeated tasks in SAS® is to write a template of the task in a macro function and then call that macro function the desired number of times. This can be difficult to implement when desiring to automate the code. Often, the number of times to execute the macro or the parameter values to use are not known until data is pulled. This paper will look at defining a macro loop from a Control Dataset, where each row (observation) is one loop iteration and each column (variable) is one local macro variable. Two implementation methods will be investigated. Both use Macro Language %do loops to repeatedly call the analysis template. The first method defines all local macro variables for all iterations a priori. Within each loop iteration, the appropriate local macro variables will be referenced with the format "&&var&i". The second method is a bit more elegant, using nested %do loops to define the local macro variables only for the current iteration. A technique for defining a Control Dataset will then be explored, which involves inner joins, outer joins, and data blocks.

Steve Ruegsegger received his Ph.D. in Electrical Engineering from University of Michigan studying control systems for semiconductor manufacturing. He then has spent 13 years at the chip factories of IBM using SAS to analyze all sorts of data – from real-time tool data to wafer

yields to bitstream fail data. He designed several SAS courses for IBM and has instructed 100's of engineers on using SAS for data analysis.

### **Building the Better Macro: Best Practices for the Design of Reliable, Effective Tools**

by Frank Dilorio

The SAS® macro language has power and flexibility. When badly implemented, however, it demonstrates a chaos-inducing capacity unrivalled by other components of the SAS System. It can generate or supplement code for practically any type of SAS application, and is an essential part of the serious programmer's tool box. Collections of macro applications and utilities can prove invaluable to an organization wanting to routinize work flow and quickly react to new programming challenges.

But the language's flexibility is also one of its implementation hazards. The syntax, while sometimes rather baroque, is reasonably straightforward and imposes relatively few spacing, documentation, and similar requirements on the programmer. In the absence of many rules imposed by the language, the result is often awkward and ineffective coding. Some amount of self-imposed structure must be used during the program design process, particularly when writing systems of interconnected applications.

This presentation will discuss a collection of macro design guidelines and coding best practices. Primarily for programmers who create systems of macro-based applications and utilities, the presentation will also be useful to programmers just starting to become familiar with the language.

---

## **BASUG Membership**

Keep your BASUG Membership up-to-date! We provide Individual Membership at \$30 per calendar year. Click [here](#) for more information or to purchase a membership.  
[Membership Signup](#).

## **Directions to Microsoft New England Research & Development Center**

Please visit the [meeting site directions page](#).

## **BASUG Contacts**

Mailing Address:

BASUG  
PO Box 170253  
Boston, MA 02117

Email [Our Webmaster](#)

# BASUG Quarterly Meeting Announcement

Learn how to get more out of your SAS analyses with this statistical double header...using propensity scores in observational research and statistical graphics with ODS. See below for details.

Please join us for these informative talks, and consider staying for the afternoon class that expands on the use of propensity score and instrumental variable methods (separate event). For more information on the afternoon training, please visit [the training announcement](#).

*Immediately following the meeting, we will provide an informal light buffet lunch for all meeting attendees. We hope you can stay for this opportunity to network and socialize with your fellow SAS users.*

- Topics:** ODS Statistical Graphics and Propensity Score Methods Using SAS
- When:** Tuesday, September 20, 2011  
8:00am – 12:15pm
- Where:** Microsoft New England Research & Development Center  
One Memorial Drive  
Conference Center First Floor  
Cambridge, MA 02142  
857-453-6000
- Directions:** Please visit the [meeting site directions page](#).
- How:** Individual, On-Line Registration Required. No Email!
- To register:** Please visit the [event registration page](#).
- Contact:** If you have questions about the meeting contact:  
[Karen Olson](#)  
[Bridget Neville](#)

## Agenda **NOTE: Earlier start and later end times**

- 8:00 Sign in and Refreshments**
- 8:30 Announcements  **Note Earlier Start Time!**
- 8:45 Creating Statistical Graphics with ODS in SAS®  
by Warren Kuhfeld, SAS Institute
- 11:00 Break
- 11:15 Propensity Score Methods Using SAS  
by R. Scott Leslie, MedImpact Healthcare Systems, Inc.
- 12:15 Meeting Adjourned – Networking Lunch  **Note Later End Time!**

*\*Note: Times (and sequence) are approximate and subject to change. Please re-visit [the BASUG website](#) for updated information.*

## Abstracts and Bios

### **Creating Statistical Graphics with ODS in SAS®**

by Warren Kuhfeld

Effective graphics are indispensable in modern statistical analysis. SAS 9.2 provides ODS Graphics, a new functionality used by statistical procedures to create statistical graphics as automatically as they create tables. ODS Graphics is also used by new procedures that are designed for graphical exploration of data. You will learn how to:

- Request graphs created by statistical procedures.
- Use the new SGPLOT, SGPANEL, SGSCATTER, and SGRENDER procedures to create customized graphs.
- Access and manage your graphs for inclusion in Web pages, papers, and presentations.
- Modify graph styles (colors, fonts, and general appearance).
- Make immediate changes to your graphs using a point-and-click editor.
- Make permanent changes to your graphs with template changes.
- Specify other options related to ODS Graphics

Most of the tutorial applies equally to SAS 9.2 and SAS 9.3. A few new features of SAS 9.3 are highlighted.

Warren Kuhfeld is manager of SAS's Multivariate Models R&D group. He received his Ph.D. in psychometrics from UNC Chapel Hill in 1985 and joined SAS in 1987. He has used SAS since 1979 and has developed SAS procedures since 1984. Warren wrote the SAS/STAT Chapter, "Statistical Graphics Using ODS" and the new SAS Press book "The Graph Template Language and the Statistical Graphics Procedures -- An Example-Driven Introduction."

### **Propensity Score Methods Using SAS**

by R. Scott Leslie

The recent push for patient-centered outcomes research has led to an increase in the use of propensity score methods in observational research to estimate treatment effects. Although observational studies can answer many relevant questions in "real world" conditions, studies that lack randomization of subjects into treatment groups must address confounding and treatment selection bias to properly estimate the effect of treatment as non-randomized groups usually differ on observed and unobserved characteristics. That is, the observed treatment effect may be due to the treatment itself or due to the differential selection into treatment groups from non-randomization.

Propensity score methods are often used to reduce confounding and treatment selection bias by mimicking randomization. Conventional regression adjustment, matching, and stratification using propensity scores are widely used techniques to adjust for treatment selection bias by balancing groups, usually a treatment group and non-treatment group, on observed characteristics. Included in this presentation is a description of these propensity score methods, an explanation of the advantages and disadvantages of each method, and applications of methods by showing examples.

R. Scott Leslie, M.P.H., is a Health Outcomes Researcher for MedImpact Healthcare Systems, Inc, a national pharmacy benefits management company. Mr. Leslie has been with MedImpact for over 6 years and is responsible for designing and conducting retrospective database and clinical intervention studies. His areas of concentration include identifying strategic project opportunities, research design and methods application, statistical analysis, and technical report and manuscript writing.

Prior to joining MedImpact in 2004, Mr. Leslie worked for the outcomes research and pharmacoeconomics department of Prescription Solutions where he assessed the outcomes of targeted disease interventions and NCQA quality improvement initiatives. Mr. Leslie is a member of the International Society of Pharmacoeconomics and Outcomes Research and the American Public Health Association. His research has been presented at national clinical and scientific conferences as well as published in various peer-reviewed medical, pharmacy and managed care journals.

Mr. Leslie received his Bachelors of Science degree from the University of California, Los Angeles and Masters of Public Health in Epidemiology and Biostatistics from Loma Linda University and is currently working towards a doctorate degree in epidemiology from San Diego State University and University of California, San Diego.

---

## **BASUG Membership**

Keep your BASUG Membership up-to-date! For more information on our membership policy or to purchase a membership, visit the [BASUG membership page](#).

## **Directions to Microsoft New England Research & Development Center**

Please visit the [meeting site directions page](#).

## **BASUG Contacts**

Mailing Address:

BASUG  
PO Box 170253  
Boston, MA 02117

Email [Our Webmaster](#)

# BASUG Quarterly Meeting Announcement

We are thrilled with the response we got to our Call-for-Papers for our Q4 Coders' Corner meeting. This meeting will feature many presenters – some first-timers, some seasoned – covering a variety of topics. Please join us to learn lots of tips 'n techniques from your fellow BASUG colleagues.

*Immediately following the meeting, we will provide an informal light buffet lunch for all meeting attendees. We hope you can stay for this opportunity to network and socialize with your fellow SAS users.*

**Topic:** Coders' Corner 2011

**When:** Wednesday, November 16, 2011, 8:15am – Noon

**Where:** Microsoft New England Research & Development Center  
One Memorial Drive  
Conference Center First Floor  
Cambridge, MA 02142  
(857) 453-6000

**Directions:** Please visit the [meeting site directions page](#).

**How:** Individual, On-Line Registration Required. No Email!

**To register:** Please visit the [event registration page](#).

**Contact:** If you have questions about the meeting Email [Lori Goldman and Bruno Berszoner](#)

## Agenda\*

- 8:15 Sign in and Refreshments
- 8:45 Announcements
- 9:00 Creating small multiples with PROC SGPanel  
by James Zeitler, Harvard Business School
- 9:12 A Practical and Efficient Approach in Generating AE (Adverse Events) Reports Within  
a Clinical Study Environment  
by Jiannan Hu, Vertex Pharmaceuticals, Inc.
- 9:29 Calendar plot in SAS  
by Meena Doshi, Tufts University School of Civil and Environmental Engineering
- 9:44 Break
- 10:01 Teaching a New Dog Old Tricks - Using the EXCEPT Operator in PROC SQL and  
Generation Data Sets to Produce a Comparison Report  
By Stanley Fogleman
- 10:13 Merging using Hash Objects  
by Sandhyasree Padmanabhan, Meyer's Primary Care Institute (UMASS Med)
- 10:30 SAS for Proc SQL Die Hards

- by Tony Gargon, CIGNA
- 10:40 Break
- 11:00 To FREQ, Perchance to MEANS  
by Christopher Bost, MDRC
- 11:10 Tips from the Trenches by Christopher Bost, MDRC
- 12:00-1:00 **Informal buffet lunch** (provided by BASUG)

*\*Note: Times (and sequence) are approximate and subject to change. Please re-visit [the BASUG website](#) for updated information.*

## Abstracts and Bios

### **Creating small multiples with PROC SGPANEL**

by James Zeitler, Harvard Business School

"Small multiples" is one of the tools espoused by Edward Tufte in *The Visual Display of Quantitative Information*. I like to think that it's a way of adding additional dimensions to simple 2-dimensional plots. Creation of small multiples has long been possible in SAS, but I've always thought it was too much trouble with the usual tools available in SAS/GRAPH. So I was delighted to discover PROC SGPANEL recently, which greatly simplified the creation of a set of plots comparing incarceration rates over time across the United States.

For the past ten years James Zeitler has supported faculty research in business disciplines at Harvard Business School and held a similar position at Bentley College for fifteen years before that. He's used SAS for more than twenty-five years.

### **A Practical and Efficient Approach in Generating AE (Adverse Events) Reports Within a Clinical Study Environment**

by Jiannan Hu, Vertex Pharmaceuticals, Inc.

When a clinical trial is at the stage of submission to regulatory authorities, or at the investigation of the interim analysis, in most cases, safety analysis plays a key part in deciding whether the trial will be ongoing and drug is approved or not according to whether the drug is safe or not. Adverse event analysis is a pivotal piece in the safety analysis, and it is common in almost every trial, and every clinical study report.

The adverse events related to study drug will be summarized, as well as by severity, by relationship. They are tabulated by MedDRA system organ class (SOC) and preferred term (PT), and each subject is counted only once, and the sorting is by descending frequency of SOC and PT. And the AE analysis is different from other safety analysis and efficacy analysis. However, there are similarities among the adverse event analysis within a study. There are challenges for a statistical programmer to provide high quality AE reports with a limited time frame. One approach is to use macros to modularize repeated work, and thus save the development time. This paper will discuss this approach in detail and share the code in getting the work done in an efficient way.

Jiannan Hu has been a SAS user for over 10 years with multiple therapeutic area experience. With his statistical background and scientific research experience he has been using SAS for the CDISC and ADaM data creation TFL programming for clinical analysis and NDA in drug development. He is interested in the automation of the clinical analysis. He is a member of NESUG and BASUG (Boston Area SAS User Group).

## **Calendar plot in SAS**

by Meena Doshi

The calendar plot is useful for showing how a single quantity varies during long periods of time. It would be so fascinating to see visually how many people in Massachusetts are affected by an infectious disease each day starting from Jan 1990 through Dec 2002. With information on two variables - the date and number of cases, and with the use of colors of your choice, the calendar plot can display this information very effectively. With the application of PROC GMAP of SAS/GRAPH and some little-known, but very powerful tricks you can create unique, custom, one-of-a-kind graphics.

Meena Doshi is a Senior Analyst at Tufts University School of Civil and Environmental Engineering. Her work primarily involves mathematical modeling of infectious diseases like Influenza, salmonella. Her prior experience extends across two industries: Marketing Research and Public Health Research. She has been using SAS for over eight years.

## **Teaching a New Dog Old Tricks - Using the EXCEPT Operator in PROC SQL and Generation Data Sets to Produce a Comparison Report**

by Stanley Fogleman

It is possible to produce a comparison report on rather short notice using the EXCEPT operator in PROC SQL to list only those rows that have changed since the last time the report was run. The generation data set represents the data today versus whenever the report was last run.

Stanley has a Bachelors (English Lit) and Master's Degree (Computer Info Systems) from Boston University and is a proud supporter of the BU Men's and Women's basketball teams. Stanley has used SAS since version 6.07 on a variety of platforms in a variety of industries. He has both Base and Advanced certifications from SAS Institute.

## **Merging Using Hash Objects**

by Sandy Padmanabhan

Using Hash tables, the merging of a huge dataset becomes easy. In the talk, I'm comparing it with the DATA step, PROC SQL and PROC FORMAT and conveying the idea that Hash Objects are an improvised addition to programming. In the next part, I'm showing a basic program of how to replace values for a variable from the detail table with the values from the reference/master table.

Sandy Padmanabhan is an aspiring statistician with Base SAS certification. She holds a Masters degree in Statistics and has around 5 years of professional experience using SAS and other statistical packages in a health care environment. In these roles, she has worked on clinical trials with ECOG and cohort studies at UMASS and as a data manager with one of the HMO research network members.

## **SAS for Proc SQL Die Hards**

by Anthony Gargon

SQL has become a mainstay in today's computing environment. Within the SAS system, PROC SQL has become a powerful and necessary component for extracting data from external relational databases, performing complex queries and joins of SAS tables and other functions for manipulating, transforming and managing data. For many analysts and programmers who are new to SAS, there is a reticence to use other PROCs or techniques that are more efficient, make greater use of the power of SAS, and easier to maintain as applications when developed by conventional coding methods.

This paper will contrast the use of Proc SQL with other, more traditional, techniques of processing data with SAS. Examples will include metrics compiled with the FULLSTIMER option for both Proc SQL program logic and contrasting SAS procedures with data step techniques.

Anthony Gargon is an Informatics Senior Specialist at CIGNA where he develops SAS applications for member advocacy programs. Prior to joining CIGNA he worked for various health insurance/managed care organizations and acute care hospitals. He has used SAS to analyze health care data on multiple platforms for more than 25 years. Before crossing the digital divide, he was an Epidemiologist with the Connecticut State Department of Health Services' Immunization Program and also a Peace Corps Volunteer. He received a Bachelors degree in 1975 from Kent State University and a Masters in 1977 from Southern Connecticut State College.

### **To FREQ, Perchance to MEANS**

By Christopher Bost

Should you run PROC FREQ or PROC MEANS on a variable? That is the question. This presentation will show how to use the PROC FREQ option NLEVELS to determine the number of levels of each variable and then, based on a user-supplied cutoff, run either PROC FREQ or PROC MEANS. The Output Delivery System (ODS) and macro variables created with PROC SQL are used to automate the process.

This talk will be of interest to any programmer who has ever typed seemingly endless lists of variable names on the TABLES and VAR statements in PROC FREQ and PROC MEANS.

### **Tips from the Trenches**

By Christopher Bost

This presentation will review some uncommon solutions to common problems, including:

- printing long character values
- exporting formatted values to Excel
- flagging inconsistencies across observations
- creating unique SAS data set names
- controlling the length of formatted values
- counting unique values within an observation
- determining the number of combinations of variables
- running PROC CONTENTS on all data sets in a folder

The different techniques addressed will offer something to SAS programmers of all levels.

Christopher Bost is the Director of the Research Technology Unit at MDRC where he is responsible for SAS training and support. He has used SAS for data management, analysis, and reporting since 1985 in fields including public health, criminal justice, managed care, and social policy. He has also taught graduate courses on managing data with SAS at New York University and The New School.

---

## **BASUG Membership**

Keep your BASUG Membership up-to-date! For more information on our membership policy or to purchase a membership, visit the [BASUG membership page](#).

## **Directions to Microsoft New England Research & Development Center**

Please visit the [meeting site directions page](#).

## **BASUG Contacts**

Mailing Address:

BASUG  
PO Box 170253  
Boston MA 02117

Email [Our Webmaster](#)

# BASUG Half-Day Training Announcement

SAS dictionary tables contain a wealth of information about a SAS session, describing contents of datasets and views, identifying macro variables, titles and footnotes, ODS destinations, and characteristics of external files. The tables are useful in and of themselves and are an essential for programmers who develop general-purpose tools or must ensure deliverables' compliance with standards.

Please join us for this informative training, and consider coming to our morning quarterly meeting as well (separate event). For more information on the morning meeting, please visit [the meeting announcement](#).

*BASUG is hosting an informal light buffet lunch between the morning and afternoon sessions We hope you can join us for this opportunity to network and socialize with your fellow SAS users.*

**Topic:** Using Dictionary Tables

**Summary:** Learn about SAS dictionary tables and their utility in creating applications.

*Please see detailed topics below.*

**Instructor:** Frank DiIorio, President of CodeCrafters, Inc.

**When:** June 15<sup>th</sup>, 2011  
1:30pm – 5:00pm

**Where:** **Microsoft New England Research & Development Center**  
One Memorial Drive  
Conference Center First Floor  
Cambridge, MA 02142  
(857) 453-6000  
*Please see directions below*

**Price:** \$125 (if paid by June 3<sup>rd</sup>)  
\$160 (if paid online after June 3<sup>rd</sup> or at-the-door)  
*Please see details below*

**Audience:** It is expected that the audience will be familiar with Base SAS.  
**Benefits:** Attendees will come away with the knowledge of how SAS dictionary tables are created and maintained, the relationships between the tables, ways to utilize the tables and real world examples.

**Contact:** If you have questions about the course contact  
[Rita Volya or Sara Hickson](#)

**Registration:** *Pre-registration is required. Please see detailed information below.*

---

## Course Description

SAS dictionary tables are a valuable metadata source; automatically generated tables that contain a wealth of information about a SAS session. The tables describe the contents of datasets and views, identifying macro variables, titles and footnotes, ODS destinations, and the characteristics of external files. The tables are useful in and of themselves (think “utility macros”). And they become even more valuable to programmers who must ensure deliverables’ compliance with standards. This course will be useful for all SAS programmers, and will use examples from the pharmaceutical industry to illustrate the concepts.

## ***Course Topics***

This seminar takes attendees on a tour of the more commonly used dictionary tables. It:

- Presents an overview of how the tables are created and maintained
- Illustrates the relationships among the tables
- Demonstrates different ways to view the tables’ contents
- Identifies usage quirks and “features”
- Gives examples of how they can be used for both generalized and pharma-specific applications

## ***Instructor Bio***

Frank DiIorio is President of CodeCrafters, Inc., a consulting firm specializing in pharmaceutical applications and SAS training.

A SAS programmer since 1975, he is the author of "SAS Applications Programming: A Gentle Introduction" and "Quick Start to Data Analysis with SAS." He is a frequent speaker at local, regional, and international SAS conferences; at last count, he has presented over 200 papers and workshops.

Frank is past President of the SouthEast SAS Users Group, and was co-chair of its 1994 and 1996 conferences. He continues to be active in the regional and local SAS user groups and was a co-founder of the Research Triangle CDISC Users Group.

---

## **Training Registration and Payment Instructions**

**With our new online system, there are changes to our procedures...  
Please read this ENTIRE section carefully!**

### ***1. Members only!***

This class is open to BASUG members only. This means that your \$30 individual annual dues must be paid **prior** to registering for the training. If you are not currently a member, you must become a member **before** registering for the training.

If you have any questions about your BASUG membership status, please contact our Membership Coordinator [Robert Rosofsky](#)

### ***2. Pricing***

**\$125** if paid **by Friday, June 3<sup>rd</sup>**

**\$160** if paid online after Friday, June 3<sup>rd</sup> or at-the-door.

### **3. Register**

### **INDIVIDUAL, ON-LINE REGISTRATION IS REQUIRED.**

You **must register** for this training. To register and purchase tickets for the class please visit the [event registration page](#).

**Please register early!** Seating and handouts are guaranteed only for pre-paid registrants.

### **4. Payment Methods**

**Credit Card:** We urge you to pay by credit card, using our online system. Make sure to purchase your ticket by (cutoff) to get the early-bird price.

**Check:** If you must pay by check, you must send an email to [Rita Volya or Sara Hickson](#) for further instructions. (You will still need to pre-register online – but we'll explain this in our email)

**Full payment is due by the day of the class - There will be NO EXCEPTIONS.** Payment at the door is by CHECK ONLY. We never accept cash. We currently cannot accept credit cards at the door. We do accept credit card payments through our online registration service.

### **5. Refund Policy**

To receive a refund for the training, please send an email to both of the training coordinators (see "Contact" above) by 5PM on Friday, June 3<sup>rd</sup>. **After Friday, June 3<sup>rd</sup>, we will refund your payment (less a \$10 processing fee) only if we can fill your seat with attendees at the door.** There are no refunds for BASUG membership dues.

---

## **BASUG Membership**

Keep your BASUG Membership up-to-date! For more information on our membership policy or to purchase a membership, visit the [BASUG membership page](#).

## **Directions to Microsoft New England Research & Development Center**

Please visit the [meeting site directions page](#).

## **BASUG Contacts**

Mailing Address:

BASUG  
PO Box 170253  
Boston MA 02117

Email [Our Webmaster](#)

# BASUG Half-Day Training Announcement

If you've ever wondered about Propensity Scoring or Instrumental Variable Methods... what they are, how they differ, or how you might use them, then this is the training for you! This training discusses various techniques to deal with bias in your data.

Please join us for this informative training, and consider coming to our morning quarterly meeting as well (separate event). For more information on the morning meeting, please visit [the meeting announcement](#).

*BASUG is hosting an informal light buffet lunch between the morning and afternoon sessions We hope you can join us for this opportunity to network and socialize with your fellow SAS users.*

**Topic:** The Use of Propensity Score and Instrumental Variable Methods to Adjust for Treatment Selection Bias

**Summary:** A comparison of statistical methods to reduce confounding and treatment selection in observational research

***Please see details below.***

**Instructor:** R. Scott Leslie, MedImpact Healthcare Systems, Inc

**When:** Tuesday, September 20, 2011  
1:30pm – 5:00pm

**Where:** **Microsoft New England Research & Development Center**  
One Memorial Drive  
Conference Center First Floor  
Cambridge, MA 02142  
(857) 453-6000  
***Please see directions below***

**Price:** \$125 (if paid by September 9<sup>th</sup>)  
\$160 (if paid online after September 9<sup>th</sup> or at-the-door)  
***Please see details below***

**Audience:** Intermediate level statisticians and SAS® programmers.

**Benefits:** Attendees will come away with the knowledge of some complex statistical methods used in the analysis of observational data.

**Contact:** If you have questions about the course contact:  
[Eric Hamann](#) or [Brian Saper](#)

**Registration:** ***Pre-registration is required. Please see detailed information below.***

---

## ***Course Description***

Observational research provides the ability to assess effects of treatments and interventions in large populations. Although observational studies can answer many relevant questions in “real world” conditions, studies that lack randomization of subjects into treatment groups must address residual confounding and treatment selection bias to properly estimate effects as non-randomized groups usually differ on observed and unobserved characteristics.

Propensity score methods are often used to reduce observed confounding and treatment selection bias by mimicking randomization. Conventional regression adjustment, matching, and stratification using propensity scores are widely used techniques to adjust for treatment selection bias by balancing groups, usually a treatment group and non-treatment group, on observed characteristics. However, controlling for unobserved confounding factors is more difficult. Instrumental variable (IV) analysis offers a way to account for unobserved factors and may have advantages over traditional techniques.

This class is intended for intermediate level statisticians and SAS® programmers.

## ***Course Topics***

Included in this half-day training class is:

- A discussion of statistical methods for analysis of observational studies
- The advantages and disadvantages of different approaches
- A comparison of results when applied to a study that evaluates the effect of therapy regimen on medication adherence
- A review of published SAS® papers on these topics

## ***Instructor Bio***

R. Scott Leslie, M.P.H., is a Health Outcomes Researcher for MedImpact Healthcare Systems, Inc, a national pharmacy benefits management company. Mr. Leslie has been with MedImpact for over 6 years and is responsible for designing and conducting retrospective database and clinical intervention studies. His areas of concentration include identifying strategic project opportunities, research design and methods application, statistical analysis, and technical report and manuscript writing.

Prior to joining MedImpact in 2004, Mr. Leslie worked for the outcomes research and pharmacoconomics department of Prescription Solutions where he assessed the outcomes of targeted disease interventions and NCQA quality improvement initiatives. Mr. Leslie is a member of the International Society of Pharmacoconomics and Outcomes Research and the American Public Health Association. His research has been presented at national clinical and scientific conferences as well as published in various peer-reviewed medical, pharmacy and managed care journals.

Mr. Leslie received his Bachelors of Science degree from the University of California, Los Angeles and Masters of Public Health in Epidemiology and Biostatistics from Loma Linda

University and is currently working towards a doctorate degree in epidemiology from San Diego State University and University of California, San Diego.

---

## Training Registration and Payment Instructions

**With our new online system, there are changes to our procedures...  
Please read this ENTIRE section carefully!**

- 1. Members only!**

This class is open to BASUG members only. This means that your \$30 individual annual dues must be paid **prior** to registering for the training. If you are not currently a member, you must become a member **before** registering for the training.

If you have any questions about your BASUG membership status, please contact our Membership Coordinator [Robert Rosofsky](#)
  - 2. Pricing**

**\$125** if paid by Friday, September 9<sup>th</sup>  
**\$160** if paid online after Friday, September 9<sup>th</sup> or at-the-door.
  - 3. Register**

**INDIVIDUAL, ON-LINE REGISTRATION IS REQUIRED.**

You **must register** for this training. To register and purchase tickets for the class please visit the [event registration page](#).

**Please register early!** Seating and handouts are guaranteed only for pre-paid registrants.
  - 4. Payment Methods**

**Credit Card:** We urge you to pay by credit card, using our online system. Make sure to purchase your ticket by Friday, September 9th to get the early-bird price.

**Check:** If you must pay by check, you must send an email to [our training coordinators](#) for further instructions. (You will still need to pre-register online – but we'll explain this in our email)

**Full payment is due by the day of the class - There will be NO EXCEPTIONS.** Payment at the door is by CHECK ONLY. We never accept cash. We currently cannot accept credit cards at the door. We do accept credit card payments through our online registration service.
  - 5. Refund Policy**

To receive a refund for the training, please send an email to [our training coordinators](#) by 5PM on Friday, September 9<sup>th</sup>. **After Friday September 9th, we will refund your payment (less a \$10 processing fee) only if we can fill your seat with attendees at the door.** There are no refunds for BASUG membership dues.
-

## **BASUG Membership**

Keep your BASUG Membership up-to-date! For more information on our membership policy or to purchase a membership, visit the [BASUG membership page](#).

## **Directions to Microsoft New England Research & Development Center**

Please visit the [meeting site directions page](#).

## **BASUG Contacts**

Mailing Address:

BASUG  
PO Box 170253  
Boston MA 02117

Email [Our Webmaster](#)