

# THE STATISTICAL CONSULTANT

Section on Statistical Consulting  
Murray K. Clayton, Editor  
Spring–Summer 1998; Volume 15, No. 1–2

American Statistical Association

## **Announcement of Travel Award Competition**

The Section on Statistical Consulting is soliciting ideas for Special Contributed Paper Sessions for the 1999 Joint Statistical Meetings to be held in Baltimore, Maryland, August 8–12, 1999. A \$500 travel award will be given for the winning proposal. The proposer will be responsible for organizing the session and lining up the speakers and abstracts in the Fall of 1998.

To submit a proposal, prepare a short (not more than one page) description of the Special Contributed Session. This should include a description of the theme of the session, the types of papers or discussions to be included, and the intended audience.

Proposal submission deadline: September 1, 1998.

Send submissions to: George P. McCabe  
Department of Statistics  
Purdue University  
West Lafayette, IN 47907-1399  
email: [mccabe@stat.purdue.edu](mailto:mccabe@stat.purdue.edu)  
fax: 765-494-0558  
phone: 765-494-6047

## **ELSEWHERE IN THIS ISSUE**

- Statisticians and Data Mining
- Comments from the Section's Past and Present Chairs
- The ASA Section on Statistical Consulting Executive Committee
- Tips on Getting the Most Out of the Joint Statistics Meetings
- Activities at the 1998 Joint Statistical Meetings

## Statisticians and Data Mining

**Steve Leeds, Ph.D.**

President, TMI Associates

5 Sunset Drive

North Caldwell, NJ 07006

TmiLeeds@aol.com

One of the first and best lessons I learned as a statistical consultant was that if you want to play a vital role in a project, you need to speak the client's language, and not spend too much time discussing the idiosyncrasies or details of one analytic technique or another. I found myself in a meeting a little over 10 years ago doing the latter. When I saw the blank stares (not even nods), I knew that I'd claimed "techie" territory. This was not the right audience for that, nor the image I wanted to present. This rule seems simple to apply, but is frequently broken.

I work as an independent consultant in Database Marketing. A large portion of the requests I receive are for direct mail response models, customer value models, and anti-attrition models. My success depends on how well I can isolate those individuals who are likely to respond to a solicitation, become profitable customers, and are not likely to attrite (i.e. leave the customer base). My success also depends on how well I can explain how the model works and how it differs from what was done in the past in terms that the client can understand.

When analyzing the results of previous mailings you will typically find that not only were the mailings biased in some way, but the names that you will ultimately be provided with to rank are also biased in some way, not to mention the fact that the mailing offer may also be changing. Complaining about design violations is not going to help. The challenge is trying to build a successful model under these conditions, and trying to push for design changes in the future.

Understanding the business, how the campaigns are executed, how names are chosen,

and understanding your data are the most important analytical skills, but these are stressed, if at all, as an afterthought.

About nine years ago, when I was working at American Express, I was first exposed to Neural Network software. At that time vendors were selling the idea of Neural Networks being the panacea for all modeling issues. Primarily, the focus was on binary response models. These vendors were happy to announce that this software could be used without statisticians. (I didn't know that we were such a problem.) Using the image of the brain the presenters showed how an Artificial Neural Network (ANN) could learn important trends in your data that "traditional statistical approaches" could not determine. On subsequent visits these vendors conceded that data preparation and an understanding of the data were quite important, though they weren't clear on who would do that.

The latest trend is called Data Mining which is part of Knowledge Discovery in Databases (KDD). Depending on whom you speak to you will get a different definition of Data Mining. The working definition seems to stress an automated ("non-user driven") analytical approach using some subset from a variety of applied statistical methods, tree-based, and artificial neural network methods to extract patterns in your database. The methods that are chosen are determined by the project analytical objective.

In fact, all of the techniques that are part of this tool kit are not new, but have been automated in a more user-friendly fashion.

Over the last couple of months I've attended some talks on data mining as well as trying to keep pace with the most recently

published books in the area. In one presentation the speaker indicated that logistic regression was an obsolete method for binary response modeling. He was recommending data mining (in his case, that meant Neural Networks). After the presentation, he privately joked to me that he presented a logistic regression model analysis to a client, but told the client it was actually a Neural Network algorithm, because he knew the client would be more impressed with that.

In another presentation describing the cons of the statistical approach, the speaker's hand-out portrayed the statistician telling a "non-statistician CEO that she or he must make a crucial business decision because of a favorable  $R^2$  value, which is usually not well received." If you've made it to the CEO's office, and you're focusing on  $R^2$ , you've violated the rule I referred to previously, in the worst place possible, and will probably not be asked back. Though, once again this is how the statistician is perceived.

While I like the addition of some new user-friendly analytical tools within the data mining paradigm, and do use a hybrid approach to a number of common modeling problems, most analysts would agree that the actual modeling or data mining exercise encompasses only about 10-20% of the overall effort, with the data preparation and understanding constituting a much more important slice of 60-80% of the effort.

What's most disturbing is the stereotypical portrayal of how the applied statistician works. The most common phrase is, "The traditional statistical approach," which essentially indicates that there is only one specific approach to the problem, every trained statistician will take the exact same path, and that this path

leaves a lot of room for improvement. In fact, each problem and dataset requires a different approach. Based on the statistician's experience with the business, product, and dynamics of both the historical and future datasets available, the approaches vary greatly. This is where the statistical consulting exercise really becomes an art form.

Another perception is that statisticians cannot do a proper analysis because all of their techniques are inflexible, as they require every variable to follow a normal distribution. This perception is not surprising, as most of the basic statistical texts spend a lot of time on the normal distribution. Even in my area some "experts" recommend transforming everything in sight to a normal distribution. This also comes under the heading of the traditional statistical approach.

Clearly, the imagery of logistic regression, discriminant analysis, and principal components pales in comparison to the biological imagery of Neural Networks and Genetic Algorithms, and the hard-hat's Data Mining for diamonds and gold in your database. This imagery is a big selling point in the client's mind, since they not only believe that they can picture the benefits, but the actual process itself.

While many companies are being wooed with the wonders of data mining as the quick fix to all of their problems, it is important as statisticians that we do a better job explaining how we function, how to effectively communicate results, and what the data-mining sales people are not telling you. Additionally, we must try to communicate in the most effective non-technical terms possible. As a consultant, doing all of this with positive results generates the most important image in the client's mind — trust.

## Some Reflections on My Year as Section Chairman

**Keith E. Muller, MA, PhD, MS**

Department of Biostatistics

McGavran-Greenberg Hall CB#7400

University of North Carolina

Chapel Hill NC 27599-7400

Keith\_Muller@UNC.EDU

### Reflection # 1: Tell Us What You Want

An ASA section is a strange animal. I mean strange-wonderful, not strange-weird. It seems to have a life of its own, seemingly impervious to lapses and ignorance by those of us trying to help it. At the same time, serving on the Executive Committee made it transparently clear that nothing happens without one or more people volunteering to work mostly for the inner satisfaction of having done something nice for others. Consequently, please tell us what you want.

At the beginning of the year, I said that section members 1) organize and help run sessions at our meetings (ASA, ENAR, WNAR), and 2) support electronic discussion groups, a web page, and this newsletter. Member of the executive committee 1) represent your interests to ASA officers and staff, and 2) serve on standing and ad hoc committees to maintain the organization. If you tell us what you want, we will pass it along, often successfully.

### Reflection #2: What Happened This Year

Think of Consulting Section members as part of the silent majority (if that does not have too much negative resonance for some of us children of the '60's). I suspect this mostly reflects the fact that our members present few contributed sessions at ASA and ENAR meetings. In contrast, we are very noisy in important ways. Our invited sessions are some of the best attended at the meetings. Furthermore, we have one of the largest section memberships. I do not think we should take either

for granted, but they reflect well on the section members. During the last year, I would categorize our work in the executive committee as focused on two goals: to increase participation and to maintain excellence. The travel award competition reflects the first goal. However, because only a fraction of members attend meetings, working towards more and electronic interactions seems even more important. Of course, if you tell us what you want, an important step in creating excellence is to actually listen and understand what you said.

### Reflection #3: Many Thanks

I wish to thank the members of the Statistical Consulting Section of ASA for allowing me the privilege of serving as Chairman of the section. I give my special thanks to the members of the Executive Committee. The American film industry awards "Oscars," and the American television industry gives "Emmys." So I've decided to give a few awards myself. Let's call them "Stat-ees" (rhymes with Emmys and ESPYs). Nobody can pronounce "statistical," anyway.

The Statee for "Most Stress Endured While Still Smiling and Doing a Great Job With Many Deadlines" goes to our past and current Program Chairs. The Statee for "Quiet Dependability, Every Organization Needs Lots of Them" goes to the newsletter editor and secretary/treasurers, past and present. The Statee for "Showing the Patience of Job, Year after Year," goes to the ASA staff for their annual task of training the newly elected (and sometimes clueless) officers of various sections. Finally, the Statee for "Best

Supporting Cast” goes to all of you members who participate in our section, whether it be at

the ASA or ENAR meetings, or electronically. Hope to see you in Dallas in August.

---

## Comments from the New Section Chair

**George P. McCabe**  
Department of Statistics  
Purdue University  
West Lafayette IN 47907-1399  
mccabe@stat.purdue.edu

These are exciting times to be a statistical consultant. With incredible computational power at our fingertips, we can put more of our energies into formulating problems, interpreting results, and drawing conclusions. A narrow definition of consulting would exclude these kinds of activities, but I would argue that we should actively involve ourselves in the subject matter. In this way we can make the transition from consultant to collaborator.

Our section charter states that we are interested in fostering the increased and improved use of statistics through effective client/consultant interchanges. One of the major ways that we do this is through the pro-

gram at the Joint Statistical Meetings, to be held August 9–13 in Dallas this year. I hope that you can attend and participate in our section business meeting and mixer on Tuesday August 11 from 6:00 to 7:30 pm.

One problem that we have is our budget. We have been carrying forward a fairly large amount from year to year and we need to explore ways in which these funds can be used to meet the needs of the members of the section. If you have some ideas about this please send them to the executive committee at [consexec@stat.purdue.edu](mailto:consexec@stat.purdue.edu) or to me ([mccabe@stat.purdue.edu](mailto:mccabe@stat.purdue.edu)). I look forward to seeing you in Dallas.

---

## Membership of the ASA Section on Statistical Consulting Executive Committee

The results of the recent elections are in, and so not only can we list the current members of the Executive Committee, but we’re in a position to list next year’s members, as well as terms of office for all members.

### Chair

George P. McCabe  
Department of Statistics  
Purdue University  
West Lafayette, IN 47907-1399  
[mccabe@stat.purdue.edu](mailto:mccabe@stat.purdue.edu)  
Term of Office: 1/98-12/99

### Past Chair

Keith E. Muller  
University of North Carolina  
CB #7400, Biostatistics  
Chapel Hill, NC 27599-7400  
[Keith\\_Muller@UNC.EDU](mailto:Keith_Muller@UNC.EDU)  
Term of Office: 1/98-12/98

### **Chair-Elect**

Janice Derr  
Pennsylvania State University  
Department of Statistics  
323 Classroom Bldg  
University Park, PA 16802-6106  
jd9@psuvm.psu.edu  
Term of Office: 1/98-12/99

### **Chair-Elect-Elect**

Brian Yandell  
University of Wisconsin—Madison  
Department of Statistics  
1210 West Dayton Street  
Madison, WI 53706-1685  
yandell@stat.wisc.edu  
Term of Office: 1/99-12/00

### **Program Chair (for the 1998 meeting)**

Lisa M. LaVange  
Quintiles, Inc.  
P O Box 13979  
Durham, NC 27709-3979  
llavange@quintiles.com  
Term of Office: 1/98-12/98

### **Program Chair-Elect (for the 1999 meeting)**

Ronald L. Wasserstein  
Washburn University  
Academic Affairs  
Topeka, KS 66621-1110  
zzwas@acc.waucc.edu  
Term of Office: 1/98-12/99

### **Program Chair-Elect-Elect (for the 2000 meeting)**

Douglas Simpson  
University of Illinois  
Department of Statistics  
101 Illini Hall  
725 South Wright Street  
Champaign, IL 61820  
dgs@uiuc.edu  
Term of Office: 1/99-12/00

### **Newsletter Editor**

Murray K. Clayton  
University of Wisconsin—Madison  
Department of Statistics  
1210 West Dayton Street  
Madison, WI 53706-1685  
clayton@stat.wisc.edu  
Term of Office: 1/95-12/98

### **Publications Officer**

Julia T. Rushing  
Bowman Gray School of Medicine  
Department of Public Health Sciences  
Medical Center Blvd  
Winston-Salem, NC 27157-1063  
jrrushing@rc.phs.wfubmc.edu  
Term of Office: 1/97-12/98

### **Publications Officer - Elect**

Nancy G. Berman  
Harbor-UCLA Medical Center  
1000 W. Carson Street  
Torrance, CA 90509  
bernam@gcrc.humc.edu  
Term of Office: 1/99-12/00

### **Secretary/Treasurer**

Clifford Pereira  
Oregon State University  
Department of Statistics  
Kidder Hall 44  
Corvallis, OR 97331-4601  
cliff@stat.orst.edu  
Term of Office: 1/98-12/99

### **Staff Liaison**

Sue Kulesher  
American Statistical Association  
1429 Duke Street  
Alexandria, VA 22314  
Sue@amstat.org  
Term of Office: 1/98-12/99

## Executive Committee

Deborah J. Rumsey-Johnson  
Kansas State University  
Department of Statistics  
Dickens Hall  
Manhattan, KS 66506  
rumsey@silver.stat.ksu.edu  
Term of Office: 1/96-12/98

Roy N. Tamura  
Eli Lilly and Company  
Lilly Corporate Center  
Drop Code 2233  
Indianapolis, IN 46285-2235  
tamura\_roy\_n@lilly.com  
Term of Office: 1/97-12/00

Diane L. Fairclough  
AMC Cancer Research Center  
Biostatistics  
1600 Pierce Street  
Denver, CO 80214  
fairclough@amc.org  
Term of Office: 1/98-12/00

## Executive Committee-Elect

Matilde Sanchez  
Merck & Company  
PO Box 2000 RY33-404  
Rahway, NJ 07065-0900  
matilde\_sanchez@merck.com  
Term of Office: 1/99-12/01

## Section Council Representatives

Michael H. Kutner  
The Cleveland Clinic Foundation  
Biostats & Epidemiology/P88  
9500 Euclid Ave  
Cleveland, OH 44195  
mkutner@bio.ri.ccf.org  
Term of Office: 1/98-12/00

I. Elaine Allen  
MetaWorks, Inc.  
470 Atlantic Avenue  
Boston, MA 02210  
isabel@metawork.com  
Term of Office: 1/97-12/98

---

## Tips on Getting the Most out of the Joint Statistical Meetings

### Richard Browne, Ph.D.

Texas Scottish Rite Hospital For Children  
2222 Welborn St.  
Dallas, Texas 75219  
rhbrowne@airmail.net

A Joint Statistical Meeting (JSM) can be a unique opportunity to learn about new methods, find out how to better use existing methods, and to meet people who could be both personal and professional friends. Preparation is the key to having a great experience. Once you get there, things can get confusing and hectic, so the more advanced planning you have done, the more you will get out of the meeting. For those of you who have never attended a JSM, or feel that you should be getting more out of it than you are, I would offer the following tips on how to get the most out

of the experience.

### Getting the Most out of the Preliminary Program

Take a couple of hours and skim through the preliminary program in the June issue of Amstat News. Mark any presentation title that looks interesting with a highlighter. Here are some hints on how to get the most out of your first pass:

1. Pay particular attention to the Invited Sessions. Those speakers have plenty

of time to make their presentation and were invited because they have a greater-than-average chance of being articulate about the subject. A new session last year was Classroom Day. This consisted of a series of one-hour presentations by speakers on the basics of relatively new methodologies. This was a great opportunity to get an introduction to items that you “always meant to read up on someday.” Most of the sessions were standing-room-only.

2. Look for Invited Panel discussions. You will hear experts debate topical issues in statistics, which will help you better understand philosophical differences that even the experts can't agree on. You can also ask them questions, either from the floor or after the session is over. You may hear some real pearls of wisdom that you won't hear elsewhere.
3. Picking through the Contributed Session papers. Contributed papers can be valuable, but be aware that the speakers have a much more limited amount of time to present. As a result, most contributed papers are not of a tutorial nature and should be attended only if you already have a background in the topic.
4. Don't overlook the Continuing Education offerings. Self-study is a great idea, but we rarely get around to it. Look over the Continuing Education courses. Some are one- or two-day courses and cost several hundred dollars. Others are for two hours and can cost under \$50. In either case, these may be the simplest and cheapest way for you to get smarter about a topic. These courses have good handouts; the day-long courses may even include the cost of a textbook in the fee. The difference between these courses and Classroom Day is that Classroom Day has no reservations and the handouts are more superficial (what do you expect for

free???)

5. Learn while you eat: Roundtable Luncheons. If your budget will allow it, look over the roundtable luncheon list to see if there are any topics of interest to you. Every luncheon I have gone to has been very stimulating. Everyone at the table (8 to 10 persons) is there because they want to talk about a topic. It is a good way to make new friends and to try out your ideas to see if others think them sound or crazy. The cost of each luncheon is about \$25. While that is a bit high, you do get a good meal served quickly. At some hotels, the lines for lunch can be quite long.
6. Contributed Posters. Posters are a great way to spend a lot of one-on-one time with an investigator to discuss many aspects of a topic. However, investigators are only at the posters from noon to 2 p.m., so you need to schedule time for them.

## Preliminary Planning

1. Make a preliminary schedule. Using the items you highlighted in the preliminary program, sketch out a schedule of things you want to attend. Be sure to allow two or more contiguous hours for going to the vendor displays, where books and software are found. Be aware that the vendors close up shop on Wednesday afternoon, even though the JSM goes on through Thursday at noon. Also, there are major presentations most evenings, such as the President's Invited Address, which you should consider in your schedule. There will likely be timing conflicts at this point, but don't fret. Most will resolve when you see the abstract in a couple of months.
2. Picking a Hotel. I strongly recommend trying to get a room in the main hotel, even if it costs a little more than

you want to pay. Being in that hotel can save precious minutes in getting to early sessions; you'll find that the popular sessions run out of chairs quickly. If you stay elsewhere, getting to the hotel where the presentations are given can be a real pain, especially if it is very hot or raining. There will also be times when you have a lot of handouts or other materials that become heavy when carried around all day. Being able to drop them in your room can be a real blessing. Also, there will be times when things are slow, or you are very tired, and then it will be quite refreshing to go back to your room and lay down.

3. Timing your arrival. If your scheduling allows, plan to arrive at the hotel on Saturday afternoon. You avoid the crushes at registration and hotel check-ins that will occur on Sunday. Also, you usually get an airfare break by staying over Saturday night. You can also check out the locations of the meeting rooms using the map you will be given. Do it! It will save you immense frustration and aggravation in the end.

### **Planning in the Last Few Weeks before the Meeting**

1. Preview the abstracts. The ASA is endeavoring to have presentation abstracts on the Web a few weeks before the meeting. Check them out. All too often, a title that sounded great in the preliminary program is *nothing* like what you expected it to be when you read the abstract. This will help you finalize your schedule of what sessions to attend. If there are still major conflicts, don't hesitate to write or e-mail one of the presenters and say that you will be unable to attend his/her presentation and that you would like a copy of any handouts when they are available. Though you may not

get the copies until several weeks after the meetings, it does keep you from missing out altogether.

2. Get out your business cards. At various presentations, you will want to leave your name and address with the speaker so you can get a reprint or otherwise be in touch. The simplest thing is to take a wad (wad=20) of business cards with you. If you are short on business cards, print your name/address/phone number/e-mail address on a sheet of paper, make photocopies, and cut them into pieces about the size of a notecard. That way they don't get crumpled up or lost.
3. Dress Code. Dress at the meetings is casual. Don't bring a suit, unless you are interviewing for a job. Beyond that, you'll see a wide range of styles of dress among the participants. If you are speaking, you might want to dress up slightly: avoid shorts, T-shirts, and blue jeans in that case.

### **When You Get There**

1. Read the Abstracts, if you haven't already. If you did not see the abstracts on the Web beforehand, look at the book of abstracts you will get at registration.
2. Check for program changes. Look at the list of program changes and room changes. It can be pretty frustrating if you show up at the wrong room, or skip another good presentation to attend a presentation that has been canceled.
3. Getting to sessions. Get to sessions ten minutes early for sessions that are important to you. Seating can be VERY tight in sessions with hot topics.
4. Look for your alumni dinner. Look on the bulletin board for dinners that are being organized for alumni from your

school. These typically happen on Monday or Tuesday night, so keep an eye out for them, or you may miss them. They are a great way to see old friends and make new ones.

5. Find the message board. Find the place where messages are posted. In the past few years, this has been an electronic bulletin board. If someone is trying to get in contact with you, this is where you should look. Conversely, it is a good way to leave a message for someone you are trying to contact.
6. Job Hunting. If you are looking for a job,

or just want to test the waters, the JSM is a great place to be. While there will be job postings on the bulletin boards, if you are serious you will want to pay the fee to get into the Interview Area. If you don't want your current employer to know about this, you can pay the fee at the JSM registration booth.

### Afterwards

If you enjoyed hearing what someone had to say, send that person an e-mail or letter and let them know. They will appreciate it and you may find yourself a new collaborator or friend!

## Activities at the 1997 Joint Statistical Meetings

### Mixer and Business Meeting

At this year's annual meeting in Dallas, the Consulting Section is sponsoring a mixer to immediately follow the business meeting. All members and prospective members are welcome at both the business meeting and the mixer. These activities will take place on Tuesday, August 11, from 6:00 p.m. to 7:30 p.m. See the program at the meetings for the location.

Last year's mixer was very successful, giving us all a chance to meet other members of the section. We especially welcome the applied statisticians within ASA who are looking for a section that fits their particular needs. Please join us for food and drink and to meet fellow consultants and applied statisticians with similar interests. See you in Dallas!

### The 1998 JSM Program

The entire program for the 1998 JSM can be found by visiting the JSM program web address: <http://www.amstat.org/meetings/jsm/1998> . The next several pages list sessions sponsored or co-sponsored by the ASA Section on Statistical Consulting. For the invited paper sessions, abstracts of the presentations have been included with the hope that they will be useful to you as you plan your meeting activities. Please take special note of the session *Curriculum for training future consultants* organized by Mark Payton, this year's travel award winner.

### Presentations Sponsored or Co-Sponsored by the Statistical Consulting Section

#### Special Contributed Papers: APPLYING RECENT STATISTICAL METHODS IN ORAL HEALTH CONSULTING

Sunday, August 9, 2:00 p.m. – 3:50 p.m.

**Organizer and Chair:** Stuart Gansky, UCSF

(2:00) Analysis of Binary Time-Dependent Co-

variates via the Cox Regression Model.  
Joan Hilton, UCSF

(2:20) Evaluation of Between- and Within-Subject Effects for the Analysis of Clustered Data.

Brian Leroux, U. of Washington and Lloyd

Mancl, U. Of Washington

- (2:40) Evaluating Accuracy of Diagnostic Tests Involving Multiple Tests of Multiple Readers Using Random Effects Models in Latent Class.

Ming Tan, St. Jude's Children's Research Hospital; Pingping Qu, U. of Kentucky

- (3:00) Multinomial Response Models in Dental Research Applications.

Stuart Gansky, UCSF and Glenn Davies, Merck & Co., Inc.

- (3:20) A Mixed Model Repeated-Measure Analysis of Alveolar Crest Height.

Elizabeth Krall, BU Goldman School of Dental Med., Julie Chandler, Merck & Co., Inc., Raul Garcia, BU Goldman School of Dental Med., Kathryn Munoz, Merck & Co., Inc., Austin Lee, BU, Dept. of Math., and Glenn Davies, Merck & Co., Inc.

- (3:40) Discussant: Tom Ten Have, U. of Pennsylvania

#### **Invited Papers: PRACTICING STATISTICS AT THE INTERFACE OF SCIENCE AND POLICY**

Sunday, August 9, 4:00 p.m. – 5:50 p.m.

**Organizer:** Linda J. Young, University of Nebraska

**Chair:** Phillip N. Ross, US-EPA

- (4:05) Interface of Science and Policy on Salmon Recovery Efforts in the Columbia River Basin.

Lyman L. McDonald, WEST, Inc.

**Abstract:** The decline of abundance of salmon in the Columbia River Basin, public policies, and recovery efforts will be briefly reviewed. My view of the role of science in these recovery efforts will be discussed with special emphasis on the interface of statistics, other science, and the public process for making policy and funding decisions.

- (4:35) Must We Choose between Production Agriculture and the Environment?

Linda J. Young, University of Nebraska

**Abstract:** A brief history of the use of pesticides and fertilizers in production agriculture is presented. From the agricultural perspective, the benefits of these chemical inputs are reviewed. The environmental concerns that have developed from the use of chemicals are

discussed. Statistical support for both perspectives and current efforts toward integration of the two approaches are considered.

- (5:05) Using Statistics to Determine Data Adequacy for Environmental Policy Decisions under Public Scrutiny.

Elizabeth J. Kelly, Los Alamos National Laboratory, Kelly Black, Neptune and Co., and Dan Michael, Neptune and Co.

**Abstract:** This paper presents EPA's Data Quality Assessment (DQA) approach for data collection and assessment in support of environmental policy decisions. This approach uses a statistical framework, based on the specification of acceptable levels for potential decision errors, to guide data collection and subsequent data assessments. The paper describes two case studies (at two DOE facilities) where the DQA process was used to support policy decisions about potential risks to the public from exposures to contaminants. These decisions had important political consequences and were subject to intense public scrutiny.

- (5:35) Discussant: Anthony R. Olsen, US-EPA.

#### **Special Contributed Panel: OPEN DISCUSSION OF THE NEW ETHICAL GUIDELINES FOR STATISTICAL PRACTICE**

Monday, August 10, 8:30 – 10:20 a.m.

**Organizer and Chair:** John Gardenier, Natl Ctr for Health Statistics

Panelists:

David Levy, George Mason Univ.

John Bailar, U. of Chicago

John Gardenier, Natl Ctr for Health Statistics

Richard Potthoff, Retired

Chamont Wang, The College of New Jersey

#### **Invited Papers: PRACTICAL POWER ANALYSES FOR STATISTICAL CONSULTING**

Monday, August 10, 10:30 a.m. – 12:20 p.m.

**Organizer:** Deborah H. Glueck, University of Medicine and Dentistry of New Jersey

**Chair:** Keith E. Muller, University of North Carolina-Chapel Hill

- (10:35) Power and Sample Size for Clinical Trials of Survival.

Jonathan J. Shuster, University of Florida

**Abstract:** This paper will be devoted to methodology for computing sample size requirements for randomized two-arm clinical trials, whose endpoints are times until an event, such as survival or event-free survival. The methodology presumes that patients enter according to a Poisson Process, and that proportional hazards can be assumed.

Three scenarios will be considered. First, we shall derive the sample size for the situation where no sequential monitoring is utilized. This is covered in the author's book, "Practical Handbook of Sample Size Guidelines for Clinical Trials" (CRC Press 1992). Two sequential designs will utilize an O'Brien-Fleming approach. We shall derive the required properties where continuous sequential monitoring is feasible, thanks to advances in information transfer. Finally, we shall consider a group sequential approach, which is the first to directly plan for outcome-based timing of looks on the information scale, from the onset of the trial. (Lan and DeMets have looked at the robustness of changing the frequency of looks during the trial, when you have a Z-value within 20% of the O'Brien-Fleming bound. See *Biometrics* 45, pp. 1017-1020, 1989). At each look where significance has not been attained, the analysis leads to calculation of the timing of the next look, according to the closeness to the stopping barrier. If the calculated timing is higher than 100% information, the trial is stopped immediately for non-significance.

For each of these sequential methods, the maximum amount of information over and above that of a non-sequential method is remarkably small. Major gains in terms of average information are realized. Maximum sample size requirements for either of the two sequential methods can be derived by multiplying the non-sequential sample size requirement by a relative efficiency factor that depends only upon the Type I and Type II errors.

(11:00) Power in the General Linear Multivariate Model.

Deborah H. Glueck, Robert Wood Johnson Medical School

**Abstract:** The general linear multivariate model (GLMM) is used often in clinical and

epidemiologic experiments. In contrast to data analysis, power calculations depend on the distributional properties of the predictors. The conditional power is calculated for a specific realization of the predictors. We describe a general class of conditional power approximations based on the method of moments and Taylor series. We demonstrate how existing conditional power approximations fall into this scheme. The unconditional power equals the expected value of the conditional power with respect to all possible stochastic realizations of any random predictors. The quantile power is the conditional power for a "worst case" scenario. We summarize unconditional and quantile power results for GLMMs with fixed and Gaussian predictors. An example power analysis illustrates the use of quantile power approximations in study design.

(11:25) Power for Categorical Data Analysis.

Ralph G. O'Brien, Cleveland Clinic Foundation, and Gwown Shieh, Cleveland Clinic Foundation

**Abstract:** From this broad topic we will focus on a pragmatic way to determine power for likelihood ratio tests in applications of logit, log-linear, and Poisson modeling from the simple to the complex. Specifically, we will show via examples how the dominating term of a good general noncentrality approximation for generalized linear models (Self, Mauritsen, and Ohara; *Biometrics*, 1992) can be computed by forming and analyzing "exemplary" data using one's regular statistical software (as per O'Brien; SUGI-11, 1986). Powers or minimal sample sizes are then computed, tabled, and/or graphed using UnifyPow, a freeware SAS macro (<http://www.bio.ri.ccf.org/power.html>).

The handout will also cover how UnifyPow easily handles statistical planning problems for many other types of categorical data analyses.

(11:50) Sample Size and Power Calculations for Studies with Correlated Data.

Guanghan (Frank) Liu, Merck Research Laboratories

**Abstract:** Correlated data occur frequently in biomedical research and have received a good deal of attention in recent years. Examples include longitudinal studies, fam-

ily studies, and ophthalmologic studies, etc. Methods to analyze correlated data have been extensively studied in recent years. In this talk, we present methods to compute sample sizes and statistical powers for studies involving correlated observations (Liu and Liang; Biometrics, 1997). This is an important step when the investigators are planning a study to address some scientific hypotheses. The proposed method is based on generalized estimating equations (GEE). The theoretical background and some additional assumptions needed for the sample size and power calculations are highlighted. Sample size and power calculation formulas will be obtained for some special cases that are commonly seen in practice. The methods will be illustrated by simulations and examples from various clinical studies including longitudinal and cross-over trials.

**Special Contributed Papers: CURRICULUM FOR TRAINING FUTURE CONSULTANTS**

Tuesday, August 11, 8:30 a.m. – 10:20 a.m.

**Organizer and Chair:** Mark Payton, Oklahoma State U

- (8:30) Statistical Curriculum for Training Future Consultants.  
Janice Derr, US-FDA
- (8:50) Statistical Consulting Training at the University of Florida.  
Ramon Littell, U. of Florida
- (9:10) What Good Is It If Clients Won't Use It?  
Walter Stroup, U. of Nebraska-Lincoln
- (9:30) Providing Training in Statistical Consulting for Students at Southern Methodist University.  
Wayne Woodward, SMU
- (9:50) Training Future Biostatisticians for the Pharmaceutical and Biotechnology Industries.  
Joseph Massaro, Boston Univ.

**Regular Contributed Papers: SPECIAL TOPICS IN STATISTICAL CONSULTING**

Tuesday, August 11, 10:30 a.m. – 12:20 p.m.

**Chair:** Sonia Davis, Quintiles, Inc.

- (10:30) Progress and Pitfalls in Power Analysis.  
Linda Goldsmith, Univ. of Louisville

(10:45) Choosing Not to Use “Real” Statistics Software.

R.K. Elswick, Jr., Virginia Commonwealth Univ., Phillip W. Iversen, Lilly Research Laboratories, and Al. M. Best, Virginia Commonwealth Univ.

(11:00) Faculty Mentoring on a Student Consulting Project: Failure Analysis of Multi-Component Sub-Systems Comprised of a Mixture of Bernoulli Components.

Dean Fern, California State U., Elliott Nebenzahl, California State U., and Leslie Freerks, California State U.

(11:15) Statistical Consulting in an International Technical Assistance Organization: A Case Study of Challenges and Rewards.

Emelita Wong, Family Health International

(11:30) Statisticians as Collaborators with Emergency Medicine Residency Programs' Research Physicians: Resources and Attitudes.  
Nancy Fenn Buderer, St. Vincent Mercy Medical Center

(11:45) Logistic Regression Diagnostics.

Renee Jones, U. of Arizona, and Denise Roe, U. of Arizona.

**Invited Papers: THE CLIENT \* CONSULTANT INTERACTION IN THE PHARMACEUTICAL INDUSTRY**

Wednesday, August 12, 10:30 – 12:20 a.m.

**Organizer and Chair:** Michael N. Boyd, Pharmaceutical Research Associates, Inc.

(10:35) Statistical Consulting in a Major Pharmaceutical Firm.

Roger F. Liddle and Ellen McSorley, Glaxo Wellcome, Inc.

**Abstract:** The clinical statistician has a critical and well-defined role in the drug development process. Members of the internal drug development team from clinical, regulatory, planning and writing represent one set of clients. Regulatory authorities and the medical community are also clients, and interaction and communication are quite different in this setting. Opportunities to broaden the current role of the clinical statistician will be presented. Comparisons will be made to other opportunities for a consulting statistician.

Another role for the pharmaceutical statistician occurs when statistical analysis is contracted out and the consulting statistician becomes the client. Various models for contracting out statistical work will impact the client/consultant relationship, but communication is the critical success factor in all of them.

A large pharmaceutical firm also holds interesting challenges for the pre-clinical and non-clinical statistician. With problems ranging from research to manufacturing, there are numerous opportunities for the consulting statistician to add value. Examples will be presented.

(11:00) Interaction between the Client and Consultant in the Data Management, Analysis, and Reporting of Clinical Trial Data.

Roger E. Flora, Pharmaceutical Research Associates, Inc.

**Abstract:** In the pharmaceutical industry, consultants and consulting organizations are being used more and more to provide the necessary data management and statistical services for accomplishing drug development. Services that may be required by a client cover a broad spectrum, and the specific needs must be communicated. Methods of analysis and the presentation of results for a given trial or drug development program can vary and require decisions to be made at various stages of the analysis and reporting process. Specific requirements in this regard, and the level of decision-making authority to be assigned to the consultant or organization providing the services, must be clearly delineated. Finally, both parties are responsible for certain activities in the process, and careful coordination of these activities is required in order to meet timelines and to provide a well-integrated final product.

The key to a successful client-consultant interaction in accomplishing the required tasks is good communication throughout the process. This presentation will address steps that can be taken by both the client and the consultant to ensure that the interaction required for the successful completion of a project occurs. Specific experiences will be used to illustrate the ideas presented.

(11:25) An FDA Reviewer's Perspective of the Client-Consultant Interaction in the Pharmaceutical Industry.

Lisa A. Kammerman, US Food and Drug Administration

**Abstract:** From my perspective as an FDA statistical reviewer, consultants to the pharmaceutical industry are primarily university faculty, independent consultants, and members of CROs. Their familiarity with the regulatory process and requirements is wide-ranging. Consultants provide support in two general areas: (1) the regulatory process from the development of INDs through the submission of NDAs and (2) the development of statistical methodology for the design and analysis of clinical studies.

For each of these areas, I will discuss the impact of the background of the consultants and their familiarity with the regulatory process on the success, from my perspective, of the client-consultant interaction. A common theme to a successful outcome is the consultant's knowledge of the drug class, indication, and regulatory requirements. Additionally, in the case of an NDA submission, the consultant's knowledge of the database is crucial.

(11:50) **Discussant:** Gary G. Koch, University of North Carolina-Chapel Hill

### **Special Contributed Panel: CONSULTANTS' FORUM: GENERALIZED LINEAR MODELS**

Wednesday, August 12, 2:00 – 3:50 p.m.

**Organizer:** Tom Loughin, Kansas State U.

**Chair:** Lisa LaVange, Quintiles, Inc.

Panelists:

Tom Loughin, Kansas State U.

Charles S. Davis, U of Iowa

Maura Stokes, SAS Institute

### **Special Contributed Panel: USING INTERACTIVE STATISTICAL SOFTWARE IN CONSULTING**

Thursday, August 13, 8:30 – 10:20 a.m.

**Organizer:** Jennifer Kendall, Zenomind Consulting

**Chair:** James Grady, U Texas Medical Branch

Panelists:

Jennifer Kendall, Zenomind Consulting

Sandra Schlotzhauer, Consultant

Randy Anderson, PPD Pharmaco

Stewart Fossceco, Knoll Pharmaceuticals

## Notes from the Editor

- You probably noticed that this issue is a combined one — combining both the Spring and Summer issues for this year. I got a bit behind on the production cycle, and this seemed to be a good way to get caught up quickly. This issue is also bigger than usual, so you're still getting your money's worth!
- Speaking of money, some questions have come up about the business side of consulting. Consider the case of an individual consulting on their own. They sign a contract with a firm to provide statistical advice. The statistician does so, and the company follows the advice. At some point later, the firm decides that the statistician's advice was flawed, and that following the advice cost them money. At this point in time, they seek financial redress. How should the consultant protect themselves against such cases? Incorporate, so that their liability is limited? Seek insurance? How, I wonder, do other professionals, such as engineers, deal with this? If you have suggestions or comments, please forward them.
- How much should you charge? This question keeps coming up, but few consultants seem willing to talk about it. One person who did was John Shade, who quoted his rates in an article for us in the Winter 1996 issue of *The Statistical Consultant*. Just to remind you, his fees ranged from £512 to £800 per day, at a time when the pound was worth roughly 1.5 U.S. dollars. Just to round out the

math on that, that works out to about £100 per hour, or \$(US) 150 per hour.

Now a piece of advice we sometimes hear is to ask local professionals how much *they* charge for an hourly rate. Accordingly, I recently took the opportunity to find out the rates charged by an attorney friend. Depending on the circumstances, they range from \$(US) 150 to \$(US) 225.

Does that seem about right? Too high? Too low?

In quite a different setting, I was reading on a usenet group about setting prices for handcrafted objects made of wood. One writer told of trying to sell a bowl without success until he *raised* the price. It seems that buyers sometimes use price as a method for assessing value.

The other piece of advice gained from the usenet woodworking group was this: although we might be tempted to charge less initially to “gain exposure,” such a strategy might backfire. Specifically, the next time the client comes back, they'll expect to pay the same low rate you charged initially, and might be upset if you raise the rates (more than, say 10%).

Comments?

As always, you can reach me via e-mail at:  
[cclayton@stat.wisc.edu](mailto:cclayton@stat.wisc.edu)  
and by US mail at:

Department of Statistics  
University of Wisconsin—Madison  
1210 W. Dayton St.  
Madison, Wisconsin 53706-1685

*THE STATISTICAL CONSULTANT*  
Murray K. Clayton, Editor  
c/o American Statistical Association  
1429 Duke St.  
Alexandria, VA 22314-3402