International Chinese Statistical Association

泛華統計協會

www.icsa.org

Bulletin

會刊

July 2006

Features:

Highlights of 2006 Applied Statistics Symposium

Current Status of Statistical Requirements for Clinical Trials in China

Contemporary Statistical Issues on Dimension Reduction

Candidates for ICSA Officers

Meeting Announcements

Table of Contents

ICSA Bulletin, July 2006

From the Editor-in-Chief

Tzu-Cheg Kao, Ph.D.

I would like to thank the Editorial Members for helping me in publishing our Bulletin, the executive members (Yi Tsong, Jun Shao, Ivan Chan, and Weiying Yuan) for continuing support, Kao-Tai Tsai (the past Editor-in-Chief) and Sue-Jane Wang for encouragement and assistance. Special thanks to Cynthia Liu for having continued to prepare the reports and articles for publication.

Some highlights: brief biographies of candidates for 2007 President-Elect, 2007 Biometrics Section Chair, and 5 Directors of the ICSA Board; a special feature article on *Current Status of Statistical Requirements for Clinical Trials in china*; two articles on high dimension reduction with a general introduction to the topic; highlights of the 2006 Applied Statistics Symposium; and information about the 2007 Applied Statistics Symposium. In the new submission guideline (detailed in this issue), please note the deadlines are December 15, and June 15 for the upcoming January, and July issues.

Serving our members in the best possible way has been my vision for the ICSA bulletin. In order to better serve our members continuously, we need volunteers for:

- Organizing the topics of general interests of our members
- Interviewing with distinguished statisticians in academia, industry and government
- Serving new, and junior statisticians
- Sharing professional accomplishments, member news, and success stories
- Listing upcoming meeting events
- Helping solicit advertisements
- Reviewing articles and reports

If you are interested in volunteering, or if you have any suggestions at all, please contact me as soon as possible.
Best Wishes,

Tzu-Cheg Kao
Editor-in Chief, International Chinese Statistical Association (ICSA) Bulletin
Tel: 301-295-9756, Fax: 301-295-1854, E-mail: tkao@usuhs.edu

Editorial Members of ICSA Bulletin

INTERNATIONAL CHINESE STATISTICAL ASSOCIATION
泛華統計協會

Dear ICSA members,

Time runs fast, already it is at mid-term of my tour of duty as the president of ICSA. I started to realize that there is so much happening during the short six-months. I have many exciting news to report to you.

ICSA Symposiums and Conferences

First, we have an extremely successful ICSA Applied Statistics Symposium at University of Connecticut. The symposium is planned with a new planning concept. They hired a management team to coordinate with meeting details and put their efforts and energies in planning and programming. They also expand the program to include more advanced topics in applied statistics. The symposium received the full collaboration of the university authorities. The results are very exciting. The three and half day program attracted more than 300 participants. The record attendance is more than 20% of the highest of the previous years. In addition to the very successful program, the symposium provided the most exiting evening entertainments at the casino and university ball room. The program and local arrangement committees were cheered for their efforts under the leaderships of Drs. Greg Wei, Min-Hue Chen and Naitee Ting. It was such a talent pool. The 2007 Applied Statistics Symposium will be held at the Triangle Research Park at North Carolina. We are looking forward to see your participation next June! ICSA is planning way ahead of its Applied Statistics Symposiums. The ICSA board have received and approved proposals of future symposium planning up to 2009 already. What a confidence index our members have on our symposiums!

On the other side of the Pacific Ocean, ICSA is also actively involved in planning for International Conference in Taipei next summer. The announcement of the conference is posted on ICSA website already. The International Conference has enjoyed a very successful run. The Taipei conference will not be exception from the tradition. Come and join us and have a great visit in Taipei in 2007.

With the success of all the association level activities, it is almost time to consider local activities. I would encourage members in the same region to work together in planning workshops or colloquia on the topics of regional interests. ICSA will make efforts to provide proper supports in financial funding when a proposal is submitted and approved by the Board.

The ICSA Applied Statistical Journal

In 2004, Dr. JP Liu, the chair of Biometrics Section conducted a survey on what the members wanted most. An applied statistical journal is on the top of the choice. Given the tremendous success of Statistica Sinica, an applied statistics is definitely a complimenting sister journal to feed the appetite of our applied members. Chair of the Program Committee, Dr. Naitee Ting agreed to organize a committee to study the feasibility and format of such a journal. The proposal has been brought to the discussion of the Board in the June meeting and received blessing by the committee for the progress. I like to congratulate Dr. Naitee Ting and Dr. Xihong Lin for making the important progress in creating an applied statistical journal that we can call it our own.

Upgrading of ICSA.org Website

It is difficult to imaging the difficulties that we face in ICSA website upgrading. In this ever-changing world of hi-tech, there are many capable persons and vendors can provide the services that we are looking for. To identify the optimal choice to balance the cost and service is both difficult and time consuming. Fortunately, when the board was at the crossroad to make a hard and uncertain decision, a new and efficient approach surfaced. We are looking forward to the new development and by the end of the year, we will be able to enjoy some of the earlier results like online credit card payments and a more efficient membership renewal services.

Election of the President, Chair of Biometrics Section and Board Members

July is ICSA election month. Because of the dedication of our members, through all the years, ICSA has not experiences any election crisis that we need to go through serious ballot re-count. You must have received the ballot. I am very excited with the candidates on the ballot. All candidates are well qualified with their backgrounds and experiences in ICSA related activities. But your voting is very important. Please make sure that you cast your ballot for the candidates that you like to see them in the future before the deadline.

This year's JSM in Seattle is only a few weeks away. Our Executive Director, Dr. Ivan Chan, and colleagues in Seattle area are working hard on the local arrangements, including the ICSDA booth, Board and membership meetings, and an annual banquet on Wednesday evening. If you are to attend the JSM, please make sure to stop by the booth and purchase your banquet tickets as well as meet your old friends and make accountant of a few new ones too.

Like every year, at the ICSA Applied Statistics Symposium, I am amazed with the many new and young faces and felt happy about the association. I met young members and found their enthusiasm with this symposium and the association. I am really proud of where ICSA is today. I still remembered vividly the first few years of the association and the symposium; I really can't imagine what we can achieve in these 17 years. My sincere thanks to our ICSA founding members for your efforts to lay down the blueprint of this organization so that we can enjoy the benefit of it. My sincere thanks also to all our members for your continuous help and devotion. I look forward to seeing you at many of the ICSA gatherings.

Yi Tsong
President, ICSA

From the Executive Director, ICSA

Ivan S. F. Chan, Ph.D.

This year we had an extremely successful Applied Statistics Symposium held at the University of Connecticut from June 14 to 17, 2006. The conference had a strong technical program and attracted a record number of participants. With the help of the University Conference Center, the meeting was professionally run, with attention to every details of the need of the participants. I would like to extend my sincere thanks to the symposium committee (chaired by Greg Wei and Ming-Hui Chen) for their tremendous efforts in making this meeting a wonderful experience.

This year we will elect several officers, including 2007 President-Elect, 2007 Biometrics Section Chair, and 5 Directors of the ICSA Board (2007-2009 term). I would like to thank the Nomination Committee, under the leadership of Xuming He, in selecting a list of very strong candidates. By now you would have already received the ballot (if not, please check the web site or contact me), and I urge all of you to participate in this important event and cast your vote. Your input is critical in selecting the future leaders of ICSA. The results of the ballot will be announced at the Annual Members meeting at the JSM in Seattle on August 9, 2006.

If you are planning to attend the Joint Statistical Meetings (JSM) in Seattle, Washington (August 6-10), be sure to visit the ICSA booth to find out what is new in ICSA and to meet new and old friends. Please also plan to attend the Annual Members Meeting on August 9 (Wednesday, 5:45-6:45 pm, Convention Center, CC-603). Following the Annual Members Meeting, please join us for a traditional Chinese dinner with lots of entertainments. We sincerely thank the Local Organizing Committee (chaired by Andrew Zhou) for coordinating this important activity.

Finally, I would like to ask you to please take a moment to check your membership information at the ICSA web site and make necessary changes if the information on the web is outdated. Please also provide your e-mail address if you forgot to do so previously. Having your updated e-mail addresses would allow us to disseminate information and communicate with you in a timely manner. If you do not remember your login ID or password, please contact Jun Zhao (Membership Committee Chair, e-mail: J.Zhao@organonusa.com) or me (e-mail: Ivan_Chan@Merck.Com).

I look forward to seeing you at the JSM, and I wish all of you a great summer.

Ivan S. F. Chan
Executive Director, ICSA

Ivan S. F. Chan, Ph.D. is Director of Clinical Biostatistics, Biostatistics and Research Decision Sciences, Merck Research Laboratories, and he can be reached by sending him an e-mail to Ivan_Chan@Merck.Com

Program Committee Report (June, 2006)

Naitee Ting, Ph.D., Chair

As you can see from the 2006 symposium report, we have had a very successful Applied Statistics Symposium at University of Connecticut this year. There were over 300 participants at this event – a record in the history of ICSA Applied Statistics Symposiums.

The Joint Statistical Meeting (JSM) at Seattle is rapidly approaching. Please refer to the announcement prepared by Dr. Andrew Zhou in this Bulletin. Again, the ICSA tradition is to hold ICSA Annual Members Meeting (5:45 pm - 6:45 pm, Wednesday, August 9) at CC-603, CC = Washington State Convention and Trade center. After the meeting, enjoy a banquet on the Wednesday evening. This year Andrew invited a Guzheng player to perform at our banquet. We hope all of you can come to the banquet to enjoy a nice dinner with other ICSA members.

The 2007 Applied Statistics Symposium will take place on June 3-6 at Research Triangle Area in North Carolina, co-organized by Shuyen Ho (phone: (919) 483-9879, email: shu-yen.ho@gsk.com) and Danyu Lin (phone: (919) 843-5134, email: lin@bios.unc.edu). Details of this upcoming symposium can be found in the symposium announcement of this Bulletin.

The ICSA International Conference will take place at Taipei in June 25-28, 2007. The name of this conference will be "The 2007 Taipei International Symposium and ICSA International Conference". Dr. Chi-Lun Cheng (email: clcheng@stat.sinica.edu.tw) at Academia Sinica is organizing this conference.

For more details of any of these upcoming activities, please visit www.icsa.org

Message from the Membership Committee

Jun Zhao, Ph.D.

As a member of the membership committee, it always make me feel happy that evidence shows the ICSA membership base is getting increasing and more members are involving in the ICSA activities. Along with the good news that the number of participants in this year's applied statistical symposium reached a historical record, date show that the number of current members is increasing. As congratulations go to the symposium organizers, I am also happy for our association, since the increase of the membership base is largely depends on the increase of participants in annual symposiums. Now an important task we need to be fulfilled is providing members better services and giving members more platforms to participate activities within the association. On the other hand, we also face some difficulty on keeping new members, which leads me to think about two existing issues: update member's current contact information and upgrade member's renew process. The good news is that the ICSA is working very hard on membership database and new website which will be easier to update and renew membership including credit cards payment. In the meantime, I, as a chair of the membership committee, encourage each member to update you contact information through the member only website of the www.icsa.org. Having your updated e-mail addresses along with your other information would allow the ICSA to disseminate information and communicate with you in a timely manner. As a statistician, you may have the same feeling as mine on missing data and biased data. If you do not remember your login ID or password, please contact Ivan Chan (Email: Ivan_chan@Merck.com) or me (Email: j.zhao@organonusa.com). We are willing to assist you.

Candidates for 2006 Election of ICSA Officers

President Elect - 2007
Biometrics Section Chair - 2007
Board of Directors (5 positions) - 2007 to 2009

Candidates for 2007 President Elect

WANG, Jane-Ling

[PRESENT POSITION] Professor, Department of Statistics, University of California at Davis **[FORMER POSITION]** Department Chair (1999-2003) and Assistant and Associate Professor in Statistics (1984-93), University of California at Davis. Associate Professor (1987-88), Wharton School of the University of Pennsylvania. Assistant Professor in Statistics (1982-84), University of Iowa. **[DEGREES]** Ph.D. in Statistics, University of California at Berkeley (1982), M.A. in Mathematics, University of California at Santa Barbara (1978), B.S. in Mathematics, National Taiwan University (1975). **[FIELD OF MAJOR STATISTICAL ACTIVITIES]** I am interested in combining statistical methodology with applications such as modeling and analyzing biomedical data and the study of longevity and aging. Current research areas also include: Analysis of longitudinal and functional data, Survival analysis, Dimension reduction methods, Joint modeling of longitudinal and survival data, and Computationally intensive statistical methods. **[SELECTED PUBLICATIONS]** Interdisciplinary work has appeared in *Science* (1994 and 1998), *Proceedings of the US National Academy of Science* (1997), *Journal of Gerontology* (1998, 1999, 2001, 2002, 2006), *Cornea* (2001, 2002), *Transactions of the American Ophthalmological Society* (2002), among others. A few recent publications in statistical journals include: 2006 - "Joint modelling of survival and longitudinal data: Likelihood approach revisited", *Biometrics*, in press (joint with F. Hsieh and Y. Tseng). 2006 - "Functional regression analysis and inference for longitudinal data", *Annals of Statistics*, 2873-2903 (joint with F. Yao and H. Müller).

2005 - "Functional data analysis for sparse longitudinal data", *Journal of the American Statistical Association*, 577-590 (joint with F. Yao and H. Müller). 2005 - "Joint modeling of accelerated failure time and longitudinal data", *Biometrika,* 587-603 (joint with F. Hsieh and Y. Tseng). 2005 - "Smoothing hazard rate", *Encyclopedia of Biostatistics*, 2nd Edition, 4986-4997. 2004 – "Functional Response Models", *Statistica Sinica,* 675-694 (joint with J. Chiou and H. Müller). **[ICSA ACTIVITES]** Board of Directors (two terms), Membership Committee, Publication Committee, Awards Committee (since 2002), Associate editor (two terms) and Co-Editor of Statistica Sinica (2002-2005), Lifetime member of ICSA. **[RELATED PROFESSIONAL ACTIVITIES]** Deming Lecture Committee (2004-2006), ISI Life Sciences Committee (2005+), IMS Council (2002-2005), ASA Fellow Committee (2000-2003), IMS Fellow Committee (2000-2003), Bernoulli Program Chair for 2007 ISI Session, IMS program Chair for 2003 JSM, Chair of Organizing Committee for 2002 Joint AMS-IMS-SIAM Summer Research Conference, Program chair for IMS Asian and Pacific Regional meeting (1997), NSF review panel (1999 and 2000), NIH BMRD Study Section (2006-2009). **[STATEMENT]** It is an honor to be nominated for the position of ICSA president and I appreciate the opportunity to serve our association. I have been involved with ICSA activities since my first service on the Board of Directors in 1993 and have witnessed the tremendous growth and accelerated reputation of ICSA within the international statistics community. These achievements are due to the unwavering support and dedicated service of our members. ICSA is by now a highly visible professional society, so it is a challenge to bring it to the next level. I will focus on the following: What are the most urgent issues faced by our organization? What directions do the members envision ICSA to take? What are the challenges we face? What's our role in the international communities? Also, how do we engage current members and recruit new ones? All these issues require inputs and insights from members and friends of ICSA. If elected, my role will be to assist the officers and members to tackle these issues and to locate needed resources.

To some extent the success of the ICSA is reflected in our flagship journal Statistica Sinica, the annual Applied Statistics Symposium, the ICSA International Conference, and its outstanding membership. Naturally, we will want to preserve these activities and qualities and to broaden their scope. For instance, members discussed establishment of an applied statistics journal. Given the excellent human resources and enthusiastic supporters we have, the time to launch such a high quality journal is due. Moreover, we could jointly sponsor existing ICSA meetings with other international organizations or institutes, and also initiate co-sponsorship of additional meetings. Such joint ventures not only broaden our scientific collaborations, but also help to build bonds with other professional societies. For several years, ICSA has explored the possibility to be a co-sponsor of the annual Joint Statistical Meeting. I would continue this effort and solicit help from our friends and members. Last but not least, the continuing success and growth of our association hinges upon our young colleagues, new researchers and students. It is vital that we nurture the junior researchers and provide placement service and career advice. One good platform to work towards these goals are the professional meetings, where we can set up special sessions with panelists, round table discussions or luncheons. Another possibility is to provide for additional dedicated meetings for new researchers and young professional statisticians. A third channel is through mentorship and awards. While ICSA has already implemented many of these activities, it could be more proactive by engaging many of our accomplished senior members.

Obviously, there is much on the plate of an ICSA president. I look forward to this exciting opportunity and welcome all suggestions if elected.

ZHANG, Heping

[PRESENT POSITION] Professor of Biostatistics, Child Study, and Statistics (2003-present), Yale University; Consultant for Iberica, Inc. (2004-present) and Eisai, Inc. (2006-present); Director of Yale Collaborative Center for Statistics in Science (2006-present); and Director of Research Methods Training in Mental Health (2003-present). **[FORMER POSITION]** Associate Professor, Yale University (1997-2003); Assistant Professor, Yale University (1992-1997); Consultant for RPR (1996-1998) and Aventis, Inc. (2002-2004); Visiting Professor, University of Zurich (2000); and Visiting Professor, University of Alabama, Birmingham (2005). **[DEGREES]** B.S. in Mathematics, Jiangxi Normal University, P.R. China (1982); Ph.D. in Statistics and Minor in Computer Sciences, Stanford University (1991). **[FIELD OF MAJOR STATISTICAL ACTIVITIES]** Asymptotic theory for latent variable models; Nonparametric methods for classification (trees) and regression (adaptive splines); Statistical Genetics; Bioinformatics; Data analysis of correlated data; Genetics of mental disorders and substance use; Post-market adverse event analysis. **[SELECTED PUBLICATIONS]** From 1991 to 2006, Dr. Zhang is the author or coauthor of over 120 research articles published in various journals including *Science, Proceedings of National Academy of Sciences, the Annals of Statistics, JASA, JRSS-B, Biometrika, Biometrics, American Journal of Epidemiology,* and *American Journal of Human Genetics*. His first English paper was published in our own journal *Statistica Sinica*. The topics of his publications range from asymptotic inference for models with irregularity problems; methodology and applications of classification trees and multivariate adaptive splines; methods, study design, and data analyses for genetic, genomic, and proteomic studies. Dr. Zhang co-authored two advanced monographs: *Recursive Partitioning in the Health Sciences* (Springer, 1999) and *Development of Modern Statistics and Related Topics* (World Scientific Publishers, 2003). **[ICSA ACTIVITIES]** Membership Committee (2004-2006); Board of Directors of ICSA (2001-2004); Executive Organization Committee of 2006 ICSA Applied Statistical Symposium. **[RELATED PROFESSIONAL ACTIVITIES]** Dr. Zhang is an associate editor for *Statistica Sinica* and *Biometrics*. He is also the editor of *Biostatistics Series* for the World Scientific Publishers. He has been a member of numerous Study Sections for the National Institutes of Health (NIH). Dr. Zhang has received many awards from the NIH including a FIRST award (1994-2000) and an Independent Scientist Award (2004-2009). He is currently the Principal Investigator on four NIH studies including an award more than $11 millions for the five-year National Genomic and Proteomic Network Study of Preterm Birth. Dr. Zhang is a member of the International Statistical Institute (1995), a Fellow of the American Statistical Association (2000), a Fellow of the Institute of Mathematical Statistics (2006), as well as a life-time member of ICSA.

[STATEMENT] I am profoundly grateful to be nominated as a candidate of president of the ICSA. In the past few years, I have been particularly fortunate to be able to attend many ICSA Applied Statistical symposiums and to witness the growth and maturity of the ICSA. Whether I am a member or a president of the ICSA, I wish to contribute to our society in

maintaining its excellent tradition by assisting junior members, expanding our basis and reach, and promoting excellent practice, training and research of statistical sciences. Many members and past presidents of the ICSA have succeeded in doing this, and I am one of the beneficiaries of their devotion, effort and wisdom. ICSA is not without challenges. Increasing membership, providing career incentives to our members and potential members, improving the quality of our professional activities such as conferences and journals, and promoting interactions among our members are no easy tasks. I congratulate and admire all past presidents for persevering in such challenges and leading us to where we are today. I will give this appointment and all the responsibilities it entails my best effort and learn as much and as quickly from those who have succeeded me in this position. Most importantly, I wish to remind anyone who is interested in my candidacy of one important point: nothing happens without the participation of our members. We should encourage our colleagues and students to join ICSA and participate in our events. Whether it is the Applied Statistics Symposium, the *Statistica Sinica*, or our world-famous parties, invite them in!

Candidates for 2007 Biometrics Section Chair

WEI, Greg

[PRESENT POSITION] Greg Wei is currently an associate director of Clinical Biostatistics of Pfizer – Groton/New London. **[FORMER POSITION]** Before he joined Pfizer he worked for Marion Merrell Dow and Sanofi-Wintrop. **[FIELDS OF MAJOR STATISTICAL ACTIVITIES]** His main expertise is in the early drug development, and he has ample experience and knowledge in clinical pharmacology studies. In the past few years, he has been a project statistician for Phase II and III studies in the areas of asthma/COPD and infectious disease. **[DEGREE]** Greg received his Ph.D. in Biostatistics from University of Wisconsin – Madison in 1989. Over the years, he has published a number of articles based on his research that has been stimulated by statistical applications in drug development. **[ICSA OFFICES & ACTIVITIES]** Greg has always been an active ICSA member. He has been an ICSA board member in the recent past. He is the chair of executive committee for ICSA Applied Statistical Symposium of 2006. Greg appreciates the opportunity to be our representative of Biopharm.

WONG, Weng Kee

[PRESENT POSITION] Professor, Department of Biostatistics, UCLA. [FORMER POSITION] Associate Professor, Department of Biostatistics, UCLA (1996-1999), Assistant Professor, Department of Biostatistics, UCLA (1990-1996). **[DEGREE]** Ph. D in Statistics, 1990, University of Minnesota; MS in Statistics, 1989, University of Minnesota; MS in Mathematics, University of Wisconsin, 1985; BSc Hons in Mathematics, National University of Singapore, 1983. **[FIELDS OF MAJOR STATISTICAL ACTIVITIES]** Optimal Design of Experiments, Biostatistics, Analysis of Cancer Trials, and, Analysis of Arthritis Rheumatoid and Scleroderma Clinical Trials. [SELECTED PUBLICATIONS] Professor Wong has published more than 90 papers in theoretical and applied statistics journals and in medical journals, including Biometrika, Journal of Royal Statistical Society, Series B., Journal of Statistical Planning and Inference, The Annals of Statistics, Statistical Neerlandica, Psychometrika, Biometrics, Drug Information Journal, Statistics in Medicine, Journal of Theoretical Biology, Journal of Official Statistics, Journal of American Statistical Association, Statistica Sinica, Journal of Biopharmaceutical Statistics, Metrika, Annals of Institute of Mathematical Statistics, Biometrical Journal, Journal of Statistical Simulation and Computation, Statistics & Probability Letters, Scandinavian Journal of Statistics, Canadian Journal of Statistics, Communications in Statistics, Sankhya, American Journal of Physiology: Endocrinology and Metabolism, Cancer Epidemiology, Biomarkers and Prevention, Arthritis Care and Research, Arthritis and Rheumatism, Journal of Rheumatology, Seminars in Arthritis and Rheumatism, Journal of Women's Health and Topics in Clinical Nutrition. He also has publications in several refereed monographs such as Statistical Methods in Clinical Studies, edited by R. B. Agostino, John Wiley (2005) and co-edited two books: an IMS Lecture Notes-Monograph series (1998) and Applied Optimal Designs by John Wiley (2006). **[ICSA OFFICES & ACTIVITIES]** Permanent Member of ICSA; Associate Editor, Statistica Sinica [2000-2002]; Organizer and Chair, "Better Writing Skills for Success", ICSA conference, 2004; Invited speaker, ICSA conference 2004; Member, Scientific Program Committee, 2004 ICSA conference **[PROFESSIONAL SERVICES]** JSM WNAR Program Chair, 2007; Member, Fisher Lectureship Award, ASA [2003-present]; WNAR Program Director [2003-present]; Advanced Program Statistics Reader, ASA, [2006-present]; Associate Editor, Biometrics [2000-present]; Associate Editor, Communications in Statistics-Theory & Methods [2003-present]; Communications in Statistics-Simulation and Computation [2003-present]; Member, Scientific Program Committee, St. Petersburg Conference, Russia 1998, 2001, 2005 and Member of several NIH Grant Review Committees, and reviewer for NSF, NSERC and NSA grants; Member, Minority Fellowship Awards, Association of Schools of Public Health, 2006; External Examiner for Graduate Students' Theses from The Netherlands (2002), Germany (2004) and Canada (2004), and, Member of a Department Accreditation Review Committee for a statistics unit in the east coast, 2006. Keynote speaker at Conference in Applied Statistics, Salamanca, Spain, 1998 and at the 14th International IOPS Conference in The Netherlands, 2004. Professor Wong is currently Principal Investigator of two grant awards, one from NIHGMS and the other from The Scleroderma Foundation.

Candidates for Directors of the ICSA Board (2007-2009)

CHAN, Wai-Sum

[PRESENT POSITION] Professor, Department of Finance, The Chinese University of Hong Kong. **[FORMER POSITIONS]** Associate Professor, Department of Statistics & Actuarial Science, University of Hong Kong, 1998-2005; Senior Lecturer, 1994-2001; Lecturer, 1989-1994, National University of Singapore. **[DEGREES]** Ph.D. in Statistics, 1989 Temple University; M.Phil. in Statistics, 1986; B.B.A. in Accounting, 1984, Chinese University of Hong Kong. **[PROFESSIONAL QUALIFICAIONS]** Fellow, Society of Actuaries, 1995; Chartered Statistician, 1999. **[FIELDS OF MAJOR STATISTICAL ACTIVITIES]** Applied Statistics with Applications to Actuarial Science, Finance, Law and Medicine. **[PUBLICATIONS]** Professor Chan has published over 60 papers in a variety of fields, including Actuarial Science area: Annals of Actuarial Science, British Actuarial Journal, North American Actuarial Journal; Law area: Law, Probability & Risk, International Journal of Evidence & Proof, Tort Law Review; Medical area: The American Journal of Epidemiology, Statistics in Medicine, Biometrics. **[ICSA ACTIVITIES]** Permanent Member of the ICSA; Member, Organizing Committee, the 5th ICSA International Conference, Hong Kong, 2001; Membership Committee, ICSA, 2002. **[PROFESSIONAL SERVICES]** Co-editor, Statistics & Finance: An Interface, Imperial College Press; Associate Editor, Journal of Economics and Management. Co-Chair, Nonlinear Time-Series Modeling, The 3rd International Association for Statistical Computing (IASC) world conference on Computational Statistics and Data Analysis, Cyprus, October 28-31, 2005.

Li, Runze

[PRESENT POSITION] Associate Professor, Department of Statistics, The Pennsylvania State University. **[FORMER POSITION]** Assistant Professor, Department of Statistics, The Pennsylvania State University (2000-2005). **[DEGREE]** Ph.D in Statistics, University of North Carolina at Chapel Hill, 2000; M.S. in Probability & Statistics, Academic Sinica, Beijing, 1993; and B.S. in Mathematics, Beijing Normal University, 1990. **[FIELDS OF MAJOR STATISTICAL ACTIVITIES]** Variable selection, functional/longitudinal data analysis, nonparametric/semiparametric regression, and modeling computer experiments **[PUBLICATIONS]** Professor Li has published one book and over 30 papers in engineering, mathematical, probability and statistical journals, including Ann. Stat., Biometrics , Biometrika, JASA and Technometrics. **[HONORS]** NSF Career award, 2004. **[PROFESSIONAL SERVICES]** IMS program chair for ENAR 05, ASA Biometrics Section program chair for JSM07 and organizers of many invited talk sessions for various professional meetings, including ENAR, JSM and ICSA applied statistics symposium. **[ICSA ACTIVITIES]** Permanent Member of ICSA; Associate Editor, Statistica Sinica (2005-).

LIU, Aiyi

[PRESENT POSITION] Investigator, Biometry and Mathematical Statistics Branch, National Institute of Child Health and Human Development, National Institutes of Health. **[FORMER POSITION]** Assistant Professor (1999-2002) Department of Biostatistics and Biomathematics, Georgetown University Medical Center, Washington, DC. **[DEGREE]** Ph.D in Statistics, 1997, M.S. in Statistics, 1995, University of Rochester; M.S. in Statistics, 1988, B.S. in Mathematics, 1986, Department of Mathematics, University of Science and Technology of China, Hefei, Anhui, China. **[FIELDS OF MAJOR STATISTICAL ACTIVITIES]** Sequential methods for clinical trials; Linear models and multivariate data analysis. **[SELECTED PUBLICATIONS]** Dr. Liu Professor Li has published over 40 statistical methodological papers including **Liu A**. "An efficient estimation of seemingly unrelated multivariate regression models with application to growth curves analysis". *Statistica Sinica* 3: 421-434, 1993; **Liu A**. "Selection of covariates and estimation of parameters in growth curve models". *Acta Mathematica Sinica* 37: 362-372, 1994; **Liu A**, Hall WJ. "Minimum variance unbiased estimation of the drift of Brownian motion with linear stopping boundaries". *Sequential Analysis* 17: 91-107, 1998; **Liu A**, Hall WJ. "Unbiased estimation following a group sequential test". *Biometrika* 86: 71-78, 1999; **Liu A**, Boyett J, Xiong XP. "Sample size calculation for planning group sequential longitudinal trials". *Statistics in Medicine* 19: 205-220, 2000; **Liu A**, Tan M, Boyett J, Xiong XP. "Testing secondary hypotheses following sequential clinical trials". *Biometrics* 56: 640-644, 2000; **Liu A**, Hall WJ. "Unbiased estimation of secondary parameters following a sequential test". *Biometrika* 88: 895-900, 2001; **Liu A**, Shih WJ, Gehan E. "Sample size and power determination for clustered repeated measurements". *Statistics in Medicine* 21: 1787-1801, 2002; **Liu A**. "Efficient estimation of two seemingly unrelated regression equations". *Journal of Multivariate Analysis* 82: 445-456, 2002; **Liu A**, Zhang Y, Gehan E, Clarke R. "Block principal component analysis with application to gene microarray data classification". *Statistics in Medicine* 21: 3465-3474, 2002; Hall WJ, **Liu A**. "Sequential tests and estimates after overrunning based on maximum-likelihood ordering". *Biometrika* 89: 699-707, 2002; Mazumdar M, **Liu A**. "Group sequential design for comparative diagnostic accuracy studies". *Statistics in Medicine* 22: 727-739, 2003; **Liu A**, Schisterman EF, Zhu Y. "On linear combinations of biomarkers to improve diagnostic accuracy". *Statistics in Medicine* 24: 37-47, 2005; **Liu A**, Wu CQ, Yu KF, Gehan E. "Supplementary analysis of probabilities at the termination of a group sequential phase II trial." *Statistics in Medicine* 24: 1009-1027, 2005; Troendle, JF, **Liu A**, Wu CQ, Yu KF. "Sequential testing for efficiency in clinical trials with non-transient effects". *Statistics in Medicine* 24: 3239-3250, 2005; **Liu A**, Hall WJ, Yu KF, Wu CQ. "Estimation following a group sequential test for distributions in the one-parameter exponential family". *Statistica Sinica*16: 165-181, 2006; Vexler A, **Liu A,** Schisterman EF, Wu CQ. "Note on distribution-free estimation of maximum linear separation of two multivariate distributions". *Journal of Nonparametric Statistics* 18: 145-158, 2006; **Liu A**, Schisterman EF, Wu CQ. "Multistage evaluation of measurement error in a reliability study". *Biometrics*, 2006, to appear; **Liu A**, Wu CQ, Yu KF, Yuan, V. "Estimation following a multivariate group sequential test". *Journal of Multivariate Analysis*, 2006, to appear.

[PROFESSIONAL SERVICES] Member of American Statistical Association, International Biometrics Society (ENAR), International Chinese Statistical Association, Institute of Mathematical Statistics, International Statistical Institute.

[ICSA ACTIVITIES] Member, ICSA 2005 Applied Statistics Symposium Planning and Program Committee; Session organizer at both 2005 and 2006 ICSA Applied Statistics Symposium; Session chair at 2004, 2005 and 2006 ICSA Applied Statistics Symposiums.

WANG, Mey

[PRESENT POSITION] Statistical Team Leader, Division of Clinical Sciences, Center for Drug Evaluation, Taiwan **[ADJUNCT POSITION]** Adjunct Assistant Professor, Taipei Medical University, Taiwan; Adjunct Assistant Professor, National Taipei Nursing College, Taiwan; Counselor, Bureau of Health Promotion, Department of Health, Taiwan **[FORMER POSITION]** Statistical Reviewer, Division of Clinical Sciences, Center for Drug Evaluation, Taiwan; Staff, Department of Pathology, National Taiwan University Hospital **[DEGREES]** Ph.D. in Biostatistics, National Taiwan University, 1998; MPH in Biostatistics, University of California, Los Angeles, 1990; BS in Public Health, National Taiwan University, 1977 **[FIELD OF MAJOR STATISTICAL ACTIVITIES]** After joining CDE, her major statistical research and interests include design and statistical methods for bridging study; adaptive designs; and epidemiology methods **[PUBLICATIONS]** "Sample size and optimal design in stratified comparative trials to establish the equivalence of treatment effects among two ethnic groups", JBS, 2002, with Yi-Hau Chen; " Clinical relevance of ethnic factors: A simulation study", Drug Information Journal suppl., 2003, with S. T. Ou, H. D. Chern, and M. S. Lin; " The use of weighted Z-tests in medical research", JBS, 2005, with K. K. Gordon Lan, Yuhwen Soo, Cynthia Siu.

WANG, Ouhong

[PRESENT POSITION] Ouhong Wang is currently Director Biostatistics in Medical Affairs at Amgen, Thousand Oaks, CA. Prior to joining Amgen in early 2005, he worked at Eli Lilly and Company in Indianapolis for more than 10 years. **[DEGREES]** Ouhong received his Ph.D. in statistics from Iowa State University in 1994. He has an M.S. degree also from Iowa State University (1991, Statistics) and a B.Eng. degree from Tsinghua University in Beijing (1989, Computer Science and Technology). **[FIELDS OF MAJOR STATISTICAL ACTIVITIES]** His interests include confounding in non-randomized clinical studies; phase I trial designs; numerical interval analysis; and missing data problems. **[PUBLICATIONS]** Ouhong published articles in Pharmaceutical Statistics, New England Journal of Medicine, Journal of Biopharmaceutical Statistics, Statistics and Computing, etc. **[ICSA ACTIVITES AND OFFICE HELD]** Ouhong currently serves on the Membership Committee, and was Chair of the Local Committee in 2000. **[RELATED PROFESSIONAL ACTIVITIES]** Ouhong served at the ASA Central Indiana Chapter and various PhRMA committees, and reviewed papers for Communication in Statistics, and Journal of Statistical Computation and Simulation.

WANG, William Wubao

[PRESENT POSITION] Associate Director, Clinical Biostatistics, Merck Research Laboratories, PA **[FORMER POSITION]** Consultant/Biometrician/ Senior Biometrician (1998 to 2005), Merck Research Laboratories, PA; Senior Biostatistician (1997-1998), Covance Inc; Programming/Statistical Consultant (1994-1997), Astra Merck Inc **[DEGREE]** Ph.D/M.S. in Statistics, Temple University 2000, 1996; M.S. in Computational Mathematics, Jilin University, 1987. **[FIELDS OF MAJOR STATISTICAL ACTIVITIES]** Bioequivalence trials, Missing/censored data and longitudinal data analyses, Vaccine clinical trials. **[PUBLICATIONS]**), "The bootstrap procedure in individual bioequivalence," Statistics in Medicine 2000, with J. Shao and S.C. Chow; " "Correlation coefficient inference on censored bioassay Data," J. of Biopharmaceutical Statistics, 2005, with L. Li and I.S.F. Chan; 'Statistical considerations for non-inferiority/equivalence Trials in vaccine development", J. of Biopharmaceutical Statistics 2006, with D.V Mehrotra, I.S.F Chan and J.F Heyse, *and 10+ other publications in biostatistical and medical journals.* **[ICSA ACTIVITIES]** Permanent Member of ICSA Since 1998; Served on the ICSA web committee 2005; Presented and chaired session at the ICSA summer symposium 1999,2002 **[PROFESSIONAL SERVICES]** Member of the organizing/steering committee and the short course committee for the FDA/Industry statistical workshop, 2005, 2006; Member of the program committee for the annual Deming conference on applied statistics, 2003 to 2006; Organized an invited session on "novel methods in vaccine trials" at ENAR 2005. Reviewer for Biometrics, 2004 and 2005.

WANG, Yuedong

[PRESENT POSITION] Professor, University of California – Santa Barbara. **[FORMER POSITIONS]** Assistant and Associate Professor (1997-2003), University of California – Santa Barbara; Assistant Research Scientist and Assistant Professor (1994-1997), University of Michigan. **[DEGREES]** Ph.D in Statistics, University of Wisconsin - Madison, 1994; M.S. in Operations Research, Chinese Academy of Science, 1987; B.A. in Mathematics, University of Science and Technology of China, 1984. **[FIELDS OF MAJOR STATISTICAL ACTIVITIES]** Biostatistical modeling; circadian rhythm; hormone pulses; microarray data analysis; nonparametric regression methods; mixed-effects models; model selection. **[PUBLICATIONS]** Dr. Wang has published 41 refereed papers in major statistical and health sciences journals, including Biometrika, JASA, JRSSB, Annals of Statistics, Statistical Science, Statistica Sinica, Biometrics, Journal of Computational and Graphical Statistics, Statistics in

Medicine, Endocrinology, Gynecologic and Obstetric Investigation, The Annals of the New York Academy of Sciences, Journal of Applied Probability, Advances in Applied Probability. **[HONORS]** Fellow of the American Statistical Association; David P. Byar Young Investigator Award from the Biometrics Section of the American Statistical Association. **[ICSA ACTIVITIES]** Permanent Member of the ICSA; Active participant of the ICSA activities. **[RELATED PROFESSIONAL ACTIVITIES]** NSF and NIH Review Panelist.

YANG, Yaning

[PRESENT POSITION] Professor of Statistics, University of Science and Technology of China **[FORMER POSITION]** Research Assistant Professor, Lab of Statistical Genetics, Rockefeller University. **[DEGREE]** Ph.D. in Statistics, Rutgers University, 2000; M.S. in Statistics, University of Science and Technology of China, 1992; B.S. in Mathematics, University of Science and Technology of China, 1989. **[FIELDS OF MAJOR STATISTICAL ACTIVITIES]** Statistical genetics, bioinformatics, epidemiology, statistical learning. **[PUBLICATIONS]** "Marginal proportional hazards models for multiple event-time data",Biometrika,2001, with Z Ying; "Efficiency of SNP haplotype estimation from pooled DNA", PNAS, 2003, with J Zhang et al.; "Asymptotics for generalized estimating equations with large cluster size", Annals of Statistics, 2003, with M Xie; "Survival Analysis of Microarray Expression Data by Transformation Models", Computational Biology and Chemistry, 2005, with J Xu, J Ott; "Computing Asymptotic Power and Sample Size for Case-Control Genetic Association Studies in the Presence of Phenotype and/or Genotype Misclassification Errors", Statistical applications in genetics and molecular biology, 2005, with F Ji et al. **[ICSA ACTIVITIES]** Member of ICSA, Volunteer for ICSA conference **[PROFESSIONAL ACTIVITIES]** Former member of ENAR

YUAN, Weiying

[PRESENT POSITION] Dr. Yuan is currently Director, Clinical Biostatistics, Johnson and Johnson Pharmaceutical Research and Development (J&J PRD). She holds a dual position of functional manager and Global Statistical Leader (GSL) in several reproductive, diabetes and urologic product teams in the area of Internal Medicine (IM). Prior to working in the IM area, Dr. Yuan worked as a GSL in the antipsychotic and neurological product teams at the J&J PRD. She has gained broad experiences in global new drug development during her fourteen years working in the pharmaceutical companies. At J&J, She has led teams of biostatisticians and SAS programmers working on various successful worldwide filings and approvals of new indications and new drugs. **[FORMER POSITION]** Prior to J&J during her tenure at Merck, she had contributed to the successful approval of Worldwide Medical Applications (WMAs) and New Drug Applications (NDAs) for new drugs on osteoporosis, analgesics, and arthritics. **[DEGREE]** Dr. Yuan received her Ph.D. and M.S. degrees in Biostatistics at University of Michigan and her B.A. in English Literature at Beijing Normal University. **[ICSA ACTIVITIES]** She is also a permanent member of ICSA and has served as Treasurer of ICSA (2004-present) and Chair of the ICSA Financial Committee (2004-2006). Previously, she was Treasurer of the ICSA Applied Symposium Account (2001-2003) and Member of the Program Committee, Applied Statistical Symposium in 2000. **[PROFESSIONAL ACTIVITIES]** Dr. Yuan is a member of ASA and DIA.

ZHAO, Jun

[PRESENT POSITION] Global Project Statistician, Organon USA Inc., Roseland, NJ, since 2000. **[DEGREE]** Ph.D. in Statistics, Rutgers University, New Brunswick, NJ, 2001; M.A. in Statistics, Rutgers University, 1998; B.S. in Mathematical Statistics, Fudan University, Shanghai, 1987. **[FIELDS OF MAJOR STATISTICAL ACTIVITIES]** Multiple comparison procedures, Clinical trial designs, Longitudinal data analysis, Experimental designs, Statistical simulations, Interim and adaptive designs, and Distribution theory. **[PUBLICATIONS AND PRESENTATIONS]** Published papers/Give presentations in applied statistics, clinical trials and finance area. **[ICSA ACTIVITIES]** Chairing the Membership Committee, 2003-2006; Fund-raiser of the Applied Statistical Symposium, 2005; Registrar for the Applied Statistical Symposium, 2000. **[OTHER PROFESSIONAL SERVICES]** Member of ASA since 2000; Basic Life Support (BLS) for Healthcare Provider, American Heart Association (AHA), since 2004; National Registry of Emergency Medical Technician (EMT-B), since 2005.

Highlights of 2006 Applied Statistics Symposium Committee Chairs

By Greg Wei, Ph.D. and Ming-Hui Chen, Ph.D.

The ICSA 2006 Applied Statistics Symposium was held from June 14 – 17 at University of Connecticut, Storrs, Connecticut. This annual statistics symposium featured three keynote talks by Professors Xiao-Li Meng of Harvard University, James O. Berger of Duke University and SAMSI, and Terry P. Speed of the University of California at Berkeley and the Walter and Eliza Hall Institute of Medical Research in Australia, and two plenary talks given by Professors Kung-Yee Liang of the National Health Research Institutes, Taiwan, R. O. C. and Johns Hopkins University and Jun S. Liu of Harvard University. Their talks on "Life becomes more colorful when you know EM, Bayes, and Wavelets ...", "Some Recent Developments in Bayesian Model Selection", "Measuring and utilizing efficiency in quantitative real-time polymerase chain reactions", "Statistics in Actions: Misuses and Alternatives", and "Sequence information, histone acetylation, and gene expression" are a great reflection of the recent development of the modern statistics. All five talks were very interesting and well attended by the symposium participants.

Four half-day and three full-day short courses were offered: "Statistical Analysis of Financial Data" by Professor Yazhen Wang, University of Connecticut, "Contributions to Discrete Distributions" by Professor Daniel Zelterman, Yale University, "Design and Analysis of Dose Response Trials" by Dr. Naitee Ting, Pfizer Global Research Development and Dr. James MacDougall, Bristol-Myers Squibb, "Active Controlled Trials" by Dr. Yi Tsong, CDER, Food and Drug Administration, "Bayesian Methods for Survival and Longitudinal Data" by Professor Joseph G. Ibrahim, University of North Carolina and Professor Ming-Hui Chen, University of Connecticut, "Statistical Methods in Bioinformatics" by Professor Jun Liu, Harvard University, and "Pharmacokinetic-Pharmacodynamic Principles for Modeling and Simulation Based Drug Development" by Dr. Marc R. Gastonguay, Metrum Research Group, LLC. We had 71 people registered for the short courses on Wednesday, June 14, 2006, prior to the technical sessions.

In addition to the three keynote and two plenary talks, we had eight sets of nine concurrent oral presentation sessions and one set of poster session during June 15-17, 2006. Among the oral presentation sessions were 60 invited sessions and 10 contributed sessions. These included: bioinformatics, clinical trials, computational biology, longitudinal and survival data analysis, and statistical genetics were hot topics in this symposium. Among other topics were: applied Bayesian statistics, career development, change-point and financial modeling, data mining, financial econometrics, industrial statistics, statistical methods in neuroscience, statistical methods in radiology research, statisticians' roles in pharmaceutical regulatory environment, the interface between statistics and finance, tournament design in sports, wavelets, and so on.

In the Fall of 2004, the five-member Executive Committee was formed. Greg Wei, Pfizer, Inc., was the chair, and the other four committee members included Ming-Hui Chen of the University of Connecticut, Fred C. Djang of Bristol Meyers Squibb, Heping Zhang of Yale University, and

Hongyu Zhao of Yale University. There were several discussions on the hosting site of the symposium. In the Fall 2005, the executive committee chose the University of Connecticut's main campus in Storrs to host the 2006 symposium. One of the main advantages of having the symposium hosted at the university was that the conference rooms were free. Ming-Hui Chen served as the chair of the Local Organizing Committee with members including Fred C. Djang, Bristol Meyers Squibb, Lynn Kuo, University of Connecticut, Ulysses Diva, University of Connecticut, Elijah Gaioni, University of Connecticut, Naitee Ting, Pfizer, Inc., Yazhen Wang, University of Connecticut, Greg Wei, Pfizer, Inc., Heping Zhang, Yale University, Hongyu Zhao, Yale University, and Bob Seguin, University of Connecticut. Lynn Kuo and Fang Yu, University of Connecticut, were the Treasurer and Registrar and the Assistant Treasurer and Registrar. The Program Committee was chaired by Hongyu Zhao, Yale University, and the committee members included Tao Huang, Yale University, Mingxiu Hu, Millenium Pharmaceuticals, Gordon Lan, Johnson & Johnson, Jane Liang, Pfizer, Inc., Jun Liu, Harvard University, Yazhen Wang, University of Connecticut, Greg Wei, Pfizer, Inc. and Zhiliang Ying, Columbia University. The Short Course Committee consisted of Yazhen Wang (chair), University of Connecticut, Ming-Hui Chen, University of Connecticut, and Greg Wei, Pfizer, Inc. Jane Liang, Pfizer, Inc. handled submissions of all contributing poster papers. The Student Award Committee included Heping Zhang (chair), Yale University, William Pan, University of New Haven, and Hongtu Zhu, Columbia University. Fred C. Djang, Bristol Meyers Squibb, chaired the J. P. Hsu Memorial Scholarship Committee and the members included Tai-Tsang Chen, Bristol Meyers Squibb and Naitee Ting, Pfizer, Inc. The Fund Raising Committee was headed by Naitee Ting, Pfizer, Inc. with members: Fred C. Djang, Bristol Meyers Squibb, Lynn Kuo, University of Connecticut, Ta-Hsin Li, IBM Watson Research Center, Greg Wei, Pfizer, Inc., and Eric Yan, Pfizer, Inc.

We had six symposium drivers including Hsu-Chih (Simon) Cheng, Cyr Emile M'lan, Sung Duk Kim, Samiran Ghosh, Ulysses Diva, and Feng Guo, all of them from University of Connecticut. They had worked six days from June 13-June 18, 2006 to pick up or drop off guesses at the Hartford train and bus station and the Bradley International Airport. Additionally, we had 10 student volunteers from the University of Connecticut including Sonali Das, Sourish Das, John Herbst, Pengfei Li,Tyler McCormick, Jaydip Mukhopadhyay, Zoe Oemcke, Balaji Raman, Changhong Song, Yingmei Xi, and Yifang Zhao, who helped to monitor equipment in the rooms where technical sessions were held.

Ulysses Diva designed and maintained the symposium web pages over last year and a half. Bob Seguin provided photos and designed several web links and the online registration form for the symposium. Tao Huang maintained abstracts for all invited guests, and helped to set up the initial technical program. Without this group effort, we would not have had this symposium. We would like to thank all committee members and all the volunteers for their effort, time, and contributions.

In this symposium, we did quite a few new things that we had not done in previous ICSA annual symposiums. With help from Bob Seguin, we designed a very colorful and informative single-page symposium advertisement. 1500 copies of the advertisement page were printed in December of 2005. Both electronic versions and hard copies of the symposium advertisement were sent out to many universities, research institutes, various companies, and individuals.

Although we do not have data to show – it did appear that the advertisements worked and it did help to spread out the information. Secondly, we offered the symposium mixer in the evening of Wednesday June 14, 2006. The mixer was greatly enjoyed by the symposium participants. Third, we offered free transportation from the train or bus station to the airport. We received positive feedbacks from many participants regarding this. Fourth, we offered free breakfast and lunches during the entire symposium. Besides the interesting technical sessions, the participants at this symposium enjoyed the casino night on Thursday June 15. About 100 participants spent about 4 hours in Mohegan Sun and Resort Casino.

The highlight of the symposium was perhaps Dr. Henry Lee's banquet speech on the evening of June 16, 2006. Dr. Henry C. Lee is one of the world's foremost forensic scientists. Dr. Lee delivered a very exciting and entertaining speech in front of over 200 banquet attendees. Dr. Lee's speech lasted over one and half hours (twice as long as was initially scheduled) – but no one noticed. After Dr. Lee's speech, there was an after-dinner karaoke as usual. People enjoyed singing their favorite songs till 11:00pm.

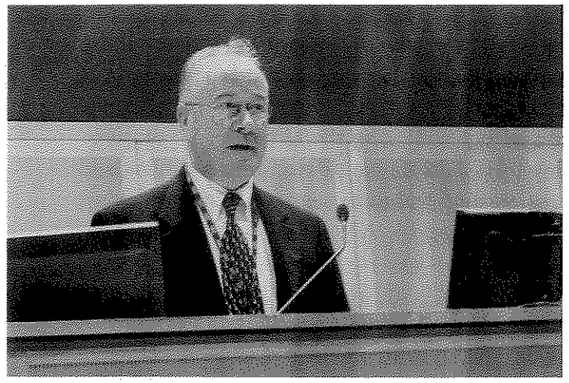
More than 318 participants from countries around the world — from US to Canada to Europe to Africa to Asia — attended the symposium. The participants for both sessions was a record high in the history of ICSA annual symposium history

This symposium received a strong support from the University of Connecticut community. The Graduate School and the College of Liberal Arts & Sciences of the University of Connecticut joining with 11 companies including Bristol-Myers Squibb Co, Boehringer Ingelheim Pharmaceuticals, Pfizer Global R&D, GSK, Organon, Amgen, Merck, IBM, Sanofi-Aventis, Johnson & Johnson, and Eisai Medical Research Inc. were the sponsors of the symposium. Dean Ross MacKinnon, Dean of the College of Liberal Arts & Sciences, and Peter J. Nichols, Provost, University of Connecticut, gave warm welcoming speeches in the opening remarks session on June15, 2006. Dipak K. Dey, Head, Department of Statistics, University of Connecticut, delivered welcoming remarks and introduced the Banquet speaker, Dr. Henry Lee, in the symposium banquet. Bob Sequin and Charity Miller of the Conference Office, University of Connecticut, had provided excellent services to the symposium.

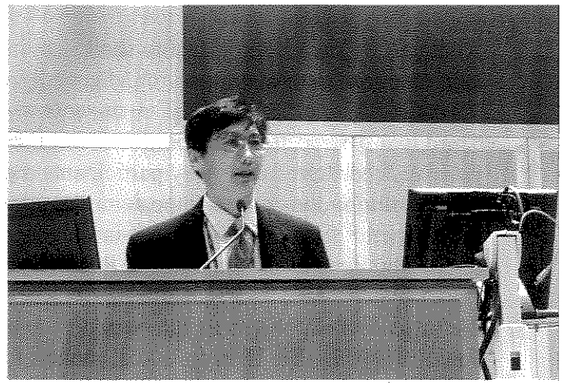
Overall, the symposium was a huge success. Many participants congratulated us for such a well organized symposium. They were impressed with our hospitality, professional services, technical sessions, the University of Connecticut's state-of-art facilities, and food. After the symposium, Xiao-Li Meng wrote to us: "It was really a great conference, one that will remain in many people's memory for long time to come. Congratulations!" Pictures about the symposium activities were shown in next few pages.

Greg Wei, Pfizer Inc., Global Research & Development, New London, Connecticut, USA. Email: greg.cg.wei@pfizer.com.

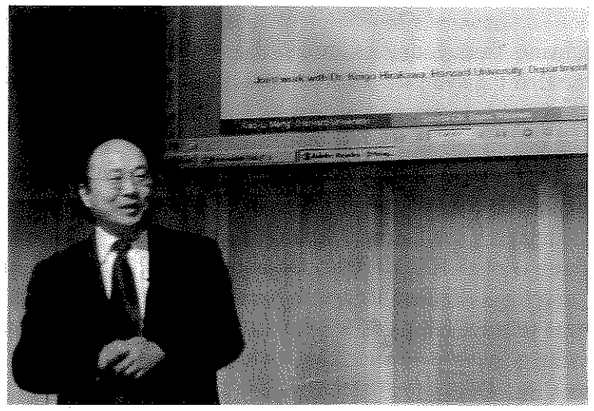
Ming-Hui Chen, Department of Statistics, University of Connecticut, Storrs, Connecticut, USA. Email: mhchen@stat.uconn.edu.



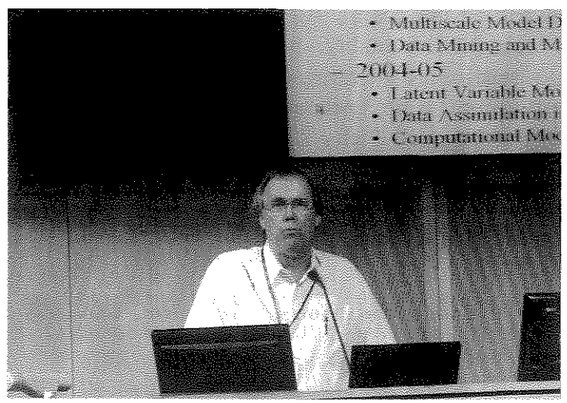
Provost Peter J. Nichols, University of Connecticut, gave welcoming remarks



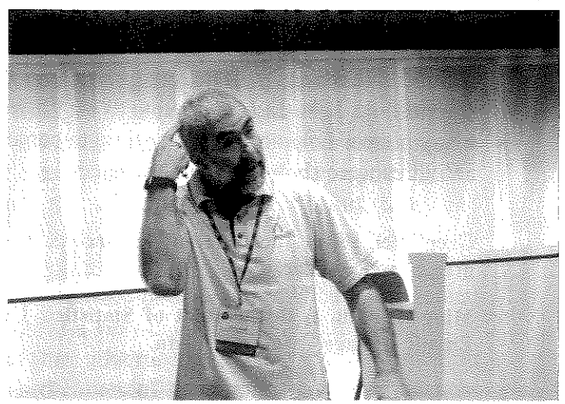
Greg Wei, Chair of the Executive Committee, gave welcoming remarks



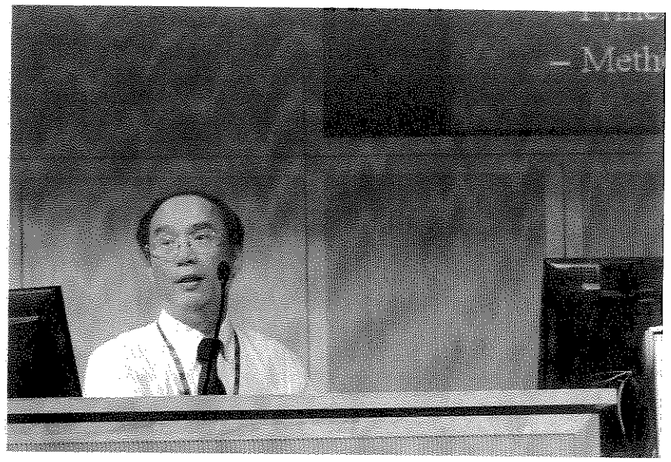
Keynote Speaker: Xiao-Li Meng, presenting "Life becomes more colorful when you know EM, Bayes, and Wavelets ..."



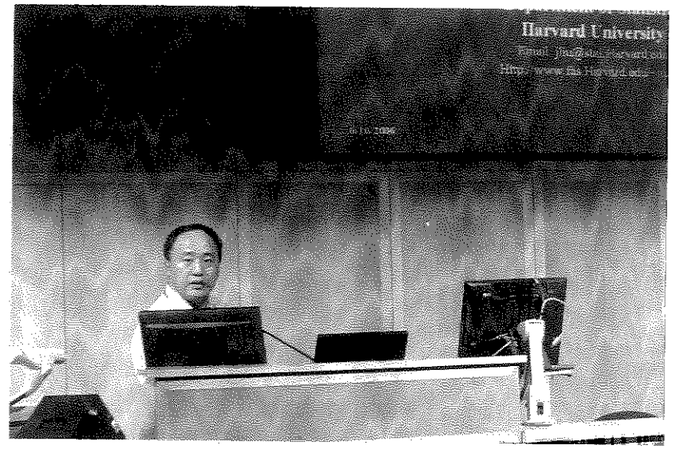
Keynote Speaker: James O. Berger, presenting "Some Recent Developments in Bayesian Model Selection"



Keynote Speaker: Terry P. Speed, presenting "Measuring and Utilizing Efficiency in Quantitative Real-time Polymerase Chain Reactions"

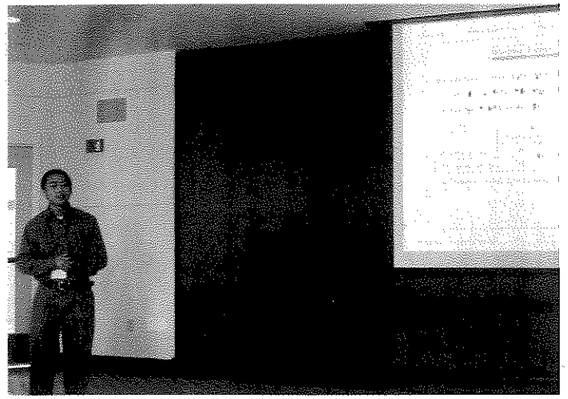


Plenary Speaker: Kung-Yee Liang, presenting "Statistics in Actions: Misuses and Alternatives"

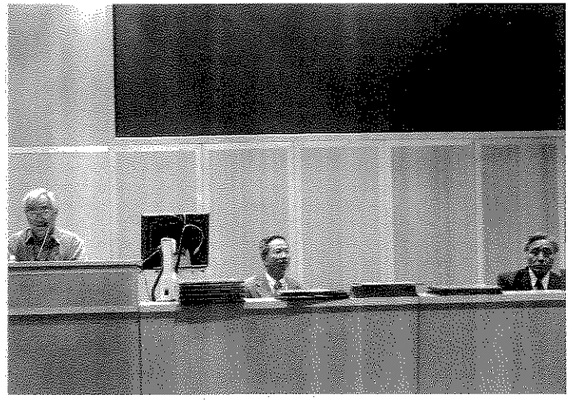


Plenary Speaker: Jun S. Liu, presenting "Sequence Information, Histone Acetylation, and Gene Expression"

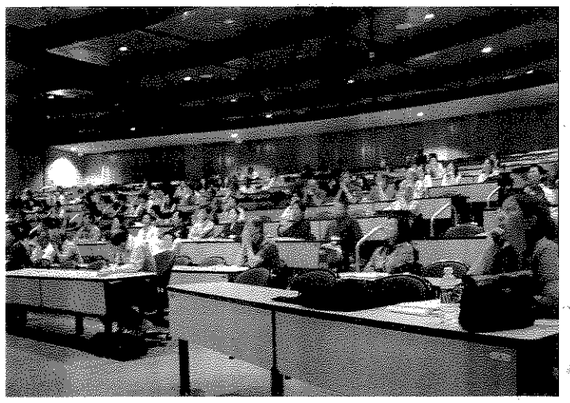
Banquet Speaker: Henry Lee, talking about "The Statistical Issues in Forensic Investigation"



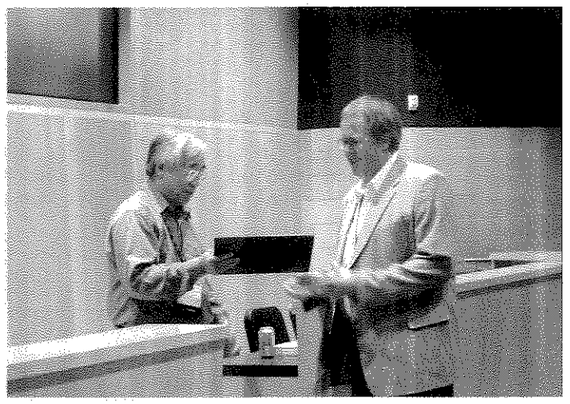
Invited Presentation



ICSA President Yi Tsong, ICSA Executive Director Ivan Chan, and ICSA Program Committee Chair Naitee Ting in the General Meeting & Awards Ceremony Session

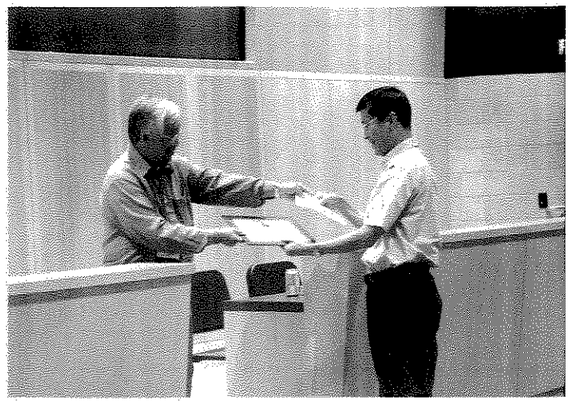


Attendees at a Keynote Session

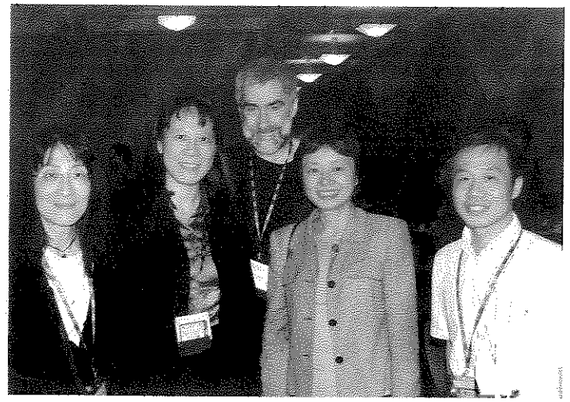


ICSA President Yi Tsong presented a plague to Keynote Speaker James O. Berger

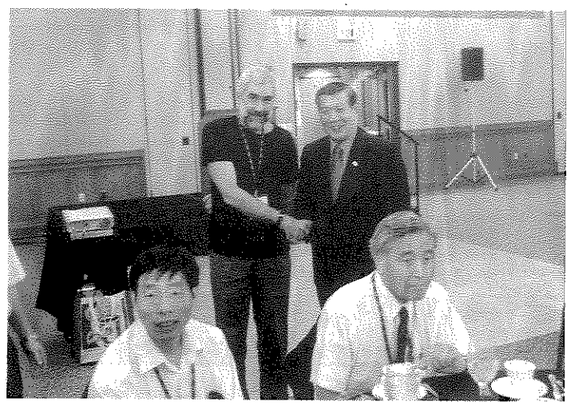
Symposium Break



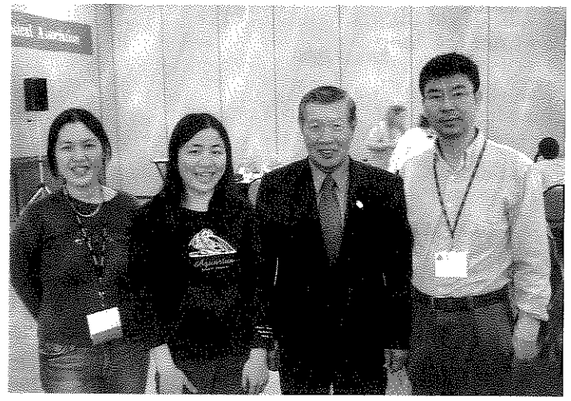
ICSA President Yi Tsong presented the certificate and a check to a student award winner



Happy Faces at the Symposium Banquet



Terry Speed and Henry Lee's Reunion at The Symposium Banquet



Henry Lee with Symposium Attendees at The Symposium Banquet I

Henry Lee with Symposium Attendees at The Symposium Banquet II

Singing a favor song at the after-dinner karaoke

Volume 16, Number 2 April 2006

Editor's Melange

Machine Learning and Data Mining

Current Status of Statistical Requirements for Clinical Trials in China

Feng Chen, Ph.D. Prof. Nanjing Medical Univeristy
Qiguang Chen, Prof. South-East Univeristy

BRIEF BACKGROUND RELEVANT TO CLINICAL TRIALS IN CHINA

China is a developing country with a huge population of around 1.27 billion in the mainland. About 1.8 million physicians are working in more than 69,000 hospitals and clinics. There are more than 6,000 pharmaceutical enterprises, research, and development institutions throughout the country. Most of factories were required to meet good manufacturing practice (GMP) standards on 30th,June of 2004, but about 10% of the enterprises failed to meet the GMP standards. The output value of gross pharmaceutical industry was more than $50 billion in 2004. From 1997 to 2002, the total health expenditure and personal health expenditure increased steadily. This huge market has attracted many multinational pharmaceutical companies and clinical contract research organizations (CROs) to invest in China in recent years.

With significant economical benefit and challenge, Chinese pharmaceutical enterprises are rapidly changing their roles from a traditional raw material suppliers and generic drug manufacturers to modern research-based ones. They strive to become major players in global pharmaceutical development by building their own niches in pharmaceutical innovation, including searching for novel therapeutic molecules, building preclinical evaluation system, and initiating extensive research and clinical collaborations with major foreign drug companies.

In China, "new drugs" refers to drugs which have not been produced previously, or to drugs which indicate, a change in the route of administration, or a change of dosage form that needs to be adopted.

Because the diagnostic and therapeutic levels vary from hospital to hospital, the hospitals participating in clinical trials for a new drug must be approved and certificated by the State Food and Drug Administration (SFDA).

The clinical trials for new drug applications are conducted in clinical centers with the

required qualification.

DEVELOPING GSP IN CHINA

The Drug Administration Law of the People's Republic of China was promulgated on July 1, 1985. Before 1998, the registration of pharmaceutical products for human use was the responsibility of the Center of Drug Evaluation, at the Ministry of Health. The Ministry of Health issued Chinese-GCP draft in 1997 after seven modifications. These guidelines were based on the ICH-GCP and FDA-GCP regulations. The State Drug Administration (SDA) was constituted in 1998, and Chinese-GCP, GMP, GLP (Good Laboratorial Practice), GAP (Good Agriculture Practice) and relevant guidelines were issued successively. Since 2001, China has formally become a member of the World Trade Organization (WTO) and has faced its greatest challenge. To fulfill its promises, the Chinese government has reviewed more than 2,300 laws and regulations relevant to the cooperation of foreign economy and trade. The central and local governments have been working hard to reform administrative procedures and to accelerate operational changes of government according to WTO standards. The Drug Administration Law of the People's Republic of China was revised and promulgated on December 1, 2001. In 2003, SFDA was established. Since then, SFDA has been working hard to draft and revise laws and regulations to meet with the international higher standards about new drug development. Chinese-GCP, GMP, GLP, GAP and relevant guidelines were revised respectively.

When the Chinese-GCP was issued, the requirements of the regulatory authorities responsible for the approval of new drugs have become more strict. The need for detailed guidance on good statistical practice in clinical research have been raised. In response to this growing need for guidance, the Drafting Committee of the Good Statistical Practice sponsored by the Center of Drug Evaluation was set up in 1997. The Drafting Committee first translated ICH E9 and FDA regulations into Chinese and introduced the regulatory requirements of the international Guidance on Statistical Principles for Clinical Trials to Chinese Statisticians. In order to ensure that the statisticians who participated in clinical research were aware of the principles of good statistical practice, two final draft versions of the Chinese *Guidance on Statistical Principles* (*GSP*) *for Clinical Trials* were completed in 1999, one for chemical and biological products, and another for traditional Chinese medications. These guidelines published in 2000 were based on the ICH-E9, E10 and FDA regulations. Dozens of training courses on GSP were held by the Drafting Committee and training center of SFDA. Thousands of biostatisticians, data managers and monitors were trained. Now, most of clinical researches adhere to these principles in conduct and reporting of clinical trials. After two years of practice, the GSPs were revised and expanded in 2003. GSP for medical instruments is currently being developed.

The quality of clinical trials in China have been greatly improved in compliance with the requirements of international GCP since the Chinese-GCP and GSP was issued.

STRUCTURE OF CHINESE-GSP

Chinese-GSP is based on ICH-E9. The contents are as follows.

1. Introduction
2. Considerations for Overall Clinical Development
   - 2.1 Exploratory trial and Confirmatory trial
   - 2.2 Variables
     - 2.2.1 Primary and Secondary Variables
     - 2.2.2 Composite Variables
     - 2.2.3 Global Assessment Variables
     - 2.2.4 Surrogate Variables
     - 2.2.5 Categorized Variables
   - 2.3 Techniques to Avoid Bias
     - 2.3.1 Randomization
     - 2.3.2 Blinding
3. Trial Design Considerations
   - 3.1. Design Configuration
     - 3.1.1. Parallel Group Design
     - 3.1.2. Crossover Design
     - 3.1.3. Factorial Designs
     - 3.1.4. Group Sequential Designs
   - 3.2. Multicenter Trial
   - 3.3. Type of Comparison
     - 3.3.1. Trials to Show Superiority
     - 3.3.2. Trials to Show Equivalence or Noninferiority
     - 3.3.3. Trials to Show Dose-Response Relationship
   - 3.4. Sample Size
   - 3.5. Data Capture and Processing
4. Trial Conduct Considerations
   - 4.1. Interim Analysis
   - 4.2. Changes in protocol
5. Data management
6. Statistical Analysis
   - 6.1. statistical analysis plan
   - 6.2. Analysis Sets
     - 6.2.1. Full Analysis Set
     - 6.2.2. Per Protocol Set
     - 6.2.3. Safety Analysis Set
   - 6.3. Missing Data and Outliers
   - 6.4. Data Transformation
   - 6.5. Evaluation of effectiveness
     - 6.5.1. Descriptive analysis
     - 6.5.2. Estimation, Confidence Intervals, and Hypothesis Testing
     - 6.5.3. Analysis of Covariates
   - 6.6. Evaluation of Safety and Tolerability
7. Statistical Reports
8. Glossary
9. Reference

SOME ISSUES IN CHINESE-GSP IMPLEMENTION

In the past decade, the quality of conducting clinical trials for drug development has been greatly improved in China. More and more statisticians, as well as data managers and programmers, who are engaged in clinical trials, adhere to the principles of ICH-E9 and Chinese-GSP in collecting, processing, analyzing, and reporting clinical trial data. But there are many aspects that should be improved. Some issues in GSP implementation are discussed as follows.

Selecting primary variable

The primary variable, also referred to as a target variable or primary endpoint, should be the variable capable of providing the most clinically relevant and convincing evidence directly related to the primary objective of the trial. In general, there is only one primary variable in a clinical trial. In phase II and Phase III trials, efficacy will usually be the primary variable, because the primary objective is to provide strong scientific evidence regarding efficacy. In phase IV, safety or tolerability may be the primary variable.

It is recommended by ICH GCP and Chinese GCP that the selection of the primary variable should reflect the accepted norms and standards in the relevant field of research. The use of a reliable and validated variable with which experience has been gained either in earlier studies or in published literature is recommended. In general, the sample size is estimated based on primary variable.

In most situations it is hard to determine the primary variable in developing Chinese traditional medicine. Even if the researcher has defined a primary variable, it is hard to measure it exactly. So, the quality of life (QOL) or other questionnaires were more used in recently years. Some questionnaires were designed for residents in western countries, but they are not always suitable for Chinese people. But researchers used these questionnaires with simple translations and changed or removed some items, without consideration for the issues of culture adaptation, item and semantic equivalence, etc. For example, item "Can you drive a car now?" was changed to "Can you ride a bike now?" in some trials since most Chinese people do not have their own cars. Item "Can you write now?" was removed because some of elder patients in rural areas were iliterate. These changes and removals were done without sufficient evidence.

Another problem with primary variables are to transform a continue variable into an ordinal rating. Chinese investigators are familiar with a scale of ordinal categorical ratings or a simple dichotomous variable when evaluating and presenting the effectiveness of a drug or a therapy even if the primary variable is a continuous one that is predefined in the protocol (e.g. change of weight, BMI, blood pressure, serum lipoprotein, etc.). They usually use ordinal ratings or a rate of occurrence as the primary variable in the place of the primary continuous one without consideration of loss of information. As emphasized in ICH E9, categorization normally implies a loss of information, consequently creating a loss of power in the analysis.

Dealing with repeated measurements

Although the observations are assumed to be independent across subjects in clinical trials, the observations that are measured repeatedly from the same subject over time are tend to be correlated. We refer to this correlation as intra-subject correlation. A few statisticians dealt with repeated measurements by means of random effect general/generalized linear models, generalized estimation equations, or multilevel models. But most of them dealt with repeated measurements on each occasion separately.

Reporting data management

Proper data management is required to ensure that a clinical trial database contains accurate, valid, and complete electronic record of the raw data, and that the database is secure. The credibility of the results of the clinical trial depend on the quality and validity of the data. It is advocated that the integration of information is more important than the mere precision of measurement. Data management, including data entry, storage, verification, correction and retrieval, is one of the most important parts of clinical research.

Software is used for data entry, details of coding systems, double data entry, blind review the database prior to finalization, and other issues are all considered. As Chinese GCP required, all steps should be documented to allow a step-by-step retrospective assessment of data quality and study performance. But few clinical trials provided the data management report.

Most of data managers use the software EpiData for data entry, storage and transformation. A few of them use FoxPro, Excel or Access.

Formulating SOP

In 1994, the Statisticians in the Pharmaceutical Industry (PSI) published the Guidelines for Standard Operating Procedures (GSOPs) on good statistical practice in clinical research. The objectives of PSI in the development of the GSOPs was to ensure that statisticians in the pharmaceutical industry were aware of the principles of good statistical practice, and to encourage adherence to these principles in the application of statistics to clinical trials, to provide guidance in the preparation of standard operating procedures, to ensure compliance with requirements of international good statistical practice, and to provide guidance in the preparation of standard operating procedures (SOP), so as to satisfy regulatory requirements with respect to the collection, process, analysis, and reporting of clinical trial data.

Most of statisticians in China adhere to the statistical principles for clinical trails. All of the works meet the regulatory requirements of statistics. But few of them write down a SOP to ensure good statistical practice in clinical research.

Applying equivalence/ noninferiority test

When an investigational product is compared to a marketed active control product, an "equivalence trial" or a "noninferiority trial" is designed to demonstrate the clinical equivalence of the test product to the marketed product, or to show that the efficacy of the test product is no worse than the active comparator.

It is recommended by ICH and FDA that the statistical analysis for equivalence or noninferiority is generally based on the use of confidence intervals. Then Type I error is appropriately controlled. For equivalence trials, equivalence is inferred when the entire confidence interval falls within the equivalence margins which were predetermined in the protocol. This is equivalent to the method of using two simultaneous one-sided tests. For noninferiority trials, a one-sided interval should be used, which is equivalent to a one-sided hypothesis test. Noninferiority is inferred if the lower confidence boundary is greater than the

lower equivalence margin which was predetermined in the protocol.

Although these are also strongly suggested by Chinese statisticians, and have been written in Chinese GSP, they have not been necessary for final approval in new drug application until now. Therefore, traditional superiority tests were wrongly conducted in equivalence trials or noninferiority trials. For example, $P>0.05$ was inferred to equivalence under the null hypothesis $\mu_1=\mu_2$. Obviously, this will increase the probability of a product which is inferior to or not equivalent to an active control product to be approved to the market.

Determining sample size

The sample size is an important issue in drafting out the protocol for a clinical trial. Both of ICH GCP and Chinese GCP require that the number of subjects in a clinical trial should be large enough to provide a reliable answer to the primary question. The method for estimating the sample size should be given in the protocol, together with the estimates of any quantities used in the calculations (such as variances, mean values, response rates, event rates, effect size , etc.,).

It is commonly recognized that the sample size should be calculated by proper statistical methods. But in Chinese GCP, there is another special requirement that the sample size in clinical trials should follow "the minimum requirement." For example, in phase II clinical trial, the patients in a test group should be at least 100; and in phase III, 300. In a clinical equivalent trial, their should be at least 100 patients in both the test and in the active control group.

This consideration is based on the safety of the patients who participated in clinical trial, for a rare side-effect is only observed in large sample sizes. But in fact, this requirement hasn't played its own role. Sample size has not been formally estimated using statistical methods for the most clinical trials, but "the least sample size" was used, especially in clinical equivalence trials and noninferiority trials.

It is dangerous to use traditional superiority tests in an equivalence/noninferiority trial with insufficient patients.

Performing interim analysis

An interim analysis is any examination of the data prior to locking the database of a clinical trial in which results (safety or efficacy) are evaluated by treatment groups. Although

the group sequential design is recommended in Chinese GSP and ICH E9, this design is not advocated by SFDA. So, few of clinical trial consider interim analysis.

CONCLUSIONS

Nowadays, the ICH guidelines on safety, efficacy, and quality are available for use by more and more countries in conducting preclinical animal toxicity studies and human clinical trials, which are highly regulated and required for new drug approval in the major pharmaceutical markets. In the past decade, clinical trials for drug development have changed dramatically in compliance with the requirements of ICH-GCP and Chinese-GCP in China. More statisticians, as well as data managers and programmers, adhere to the principles of ICH-E9 and Chinese-GSP initiatively, in order to satisfy regulatory requirements with respect to the collection, processing, analysis, and reporting of clinical trial data. However, there are many aspects that should be improved. It can be predicted hopefully that, within the next decade, clinical trials in China will meet the international requirements completely.

We would like to thank Dr. Greg G C Wei for helpful comments on the manuscript.

REFERENCE

1 National Bureau of Statistics of China. Bulletin of Fifth National Population Census (No. 1). 2002

2 http://www.stats.gov.cn/english/statisticaldata/yearlydata/

3 SFDA. Good Clinical Practice (GCP). 1998

4 SFDA. *Guidance on Statistical Principles (GSP) for Clinical Trials.* 2000

5 GW. Sang. Current Status of Clinical Trial on Drugs and GCP in China. *Drug Information Journal,* 31:1109-1125, 1997

6 Wllick Wong. Clinical Research in China. *Drug Information Journal,* 31:93-95, 1997

7 KI Kaitin. Global Drug Development and International Harmonization: The emergence of China as a world Pharmaceutical Player. *Drug Information Journal,* 32:1187S-1191S, 1998

8 K Tsutani. General View of Clinical Trials and GCP in East Asia. *Drug Information Journal,* 31: 1057-1064, 1997

9 ME Rosenberg. Implementing GCPs in Asia. *The Quality Assurance Journal.* 4:73-77 2000

10 DS Chien. The Role of Asia in Global Drug Discovery and Development. *Drug Information Journal,* 37: 3S-9S, 2003

Contemporary Statistical Issues on Dimension Reduction

Brief Introduction on Reduction of High Dimensional Data

Bing Li, Ph.D.
Professor of Statistics
Department of Statistics
Pennsylvania State University
326 Thomas Building
University Park, PA 16802
Email: bing@stat.psu.edu

In dealing with high dimensional data, the fundamental question is whether they actually lie in a low dimensional space. This has two meanings --- whether they themselves lie in a low dimensional space or whether they lie effectively in a low dimensional space in predicting their responses. The first problem is the un-supervised dimension reduction, and the second, the supervised dimension reduction. When we say "lying in a low-dimensional space" we mean it in a statistical sense. That is, what lie outside the low-dimensional space are due to randomness and specific to subjects, or, in the second case, irrelevant for predicting the response. This low-dimensional object that we seek is the "pattern", the "classifier", the "common feature", the "single index", or the "sufficient plot."

Problems of this type are increasingly common. For example, the high dimensional data set in question can be the intensities of an image at its different locations, microarrays of gene expressions, relative frequencies of different patterns (or "words") in a DNA sequence or a webpage. Because of these demands we have seen a surge of research activities in this area. The following two articles reflect these researches from different perspectives. The paper by Hu and Xu focuses on the use of Singular Value Decomposition in analyzing microarrays, which is in the flavor of un-supervised dimension reduction, and the article by Li gives an overview of recent developments in supervised dimension reduction.

Editor's notes: I would like to give thanks to Jun Shao for organizing the Contemporary Statistical Issues on Dimension Reduction, Jianhua Hu and Xuming He for applying singular value decompositions to Microarray data, and Bing Li for briefing on reduction of high dimensional data, and overview of issues related to the sufficient dimension reduction.

Sufficient dimension reduction: an overview

BING LI
Pennsylvania State University

1 Basic ideas

Sufficient dimension reduction is about reducing the dimension of a regression relation by projecting a high dimensional data onto a low dimensional subspace. It was originally introduced as graphical device to achieve comprehensive views of the regression relation: the traditional regression plots such as scatter plot matrices and residual plots are intrinsically marginal, and as such cannot reflect the true regression relation often hidden in high dimensional spaces. But what has attracted increasing attention recently is another feature of these methods — their ability to avoid the "curse of dimensionality"; that is, the accuracy of high-dimensional smoothing decreases exponentially as the dimension of the object to be smoothed increases. The pioneering works in this area are Li and Duan (1989), Li (1991, 1992), Cook and Weisberg (1991), and Cook (1994, 1996). This area of research has gain considerable momentum recently because of the increasing demands for pre-processing high dimensional data set.

Sufficient dimension reduction avoids high dimensional smoothing — which is the source of the curse of dimensionality — by using the symmetry in the high-dimensional predictors, and this symmetry arises naturally in the lower-dimensional projection of high dimensional data. This can be illustrated by a commonly used dimension reduction method called the Sliced Inverse Regression (SIR). Suppose that X is a p-dimensional vector and Y is a scalar, and that we are interested in studying the regression relation between Y and X, with X being the predictor and Y being the response. Suppose that Y depends on X only through a set of linear combinations of X, say $\beta^T X$, where β is a $p \times q$ matrix with $q \leq p$. The goal of sufficient dimension reduction is to find these linear combinations. For illustration, let us take $q = 1$ and X to be standardized so that $E(X) = 0$ and $\text{var}(X) = I_p$.

It can be shown that if the conditional expectation $E(X|\beta^T X)$ is a linear function of X, then the inverse conditional expectation

$$E(X|Y) \tag{1}$$

is parallel to β, regardless of the dimension of X, and regardless of the shape of the relation between X and Y. Note that the estimation of (1) only involves 1-dimensional smoothing, rather than p-dimensional smoothing if one attempts to estimate $E(Y|X)$. This is the reason that we can maintain the parametric rate (ie $\sqrt{n}$-rate) of convergence regardless of the dimension of X, thus avoiding the curse of dimensionality. Intuitively, when we smooth over Y, a neighborhood of Y can involve X's that are far apart. Thus we are in fact averaging X globally, and it is the symmetry in X that allows us to do so without causing bias.

At first glance the linear condition appears restrictive, because β is unknown. However, it is justified by a deep result regarding the projection of high-dimensional data, developed by Diaconis and Freedman (1983), and Hall and Li (1994). It states that, if the dimension p is large, then most of the lower-dimensional projections $\beta^T X$ are approximately normal. Precisely, suppose β is random vector uniformly distributed on a unit sphere, and Z is standard normal random variable. Then, for any (Borel) set A, the difference $P(\beta^T X \in A|\beta) - P(Z \in A)$ converges in β-probability to 0 as $p \to \infty$. Intuitively, projection, being a linear operation, amounts to averaging many small components when p is large, and hence results in normality. Based on our experience this normality is often approached rather quickly. For example, if we take a random vector that is uniformly distributed over a disc in $\mathbb{R}^p$, then, at $p = 10$, its marginal distribution is already very close to normal. Considering that nowadays data sets often have much larger dimensions, the linearity condition should be reasonable in many applications. Moreover, for lower dimensional X, it is possible to achieve approximate linearity conditions by Box-Cox transformation, or a re-weighting method introduced by Cook and Nachtsheim (1994).

Another important point is that while the assumption that Y depends on X only through $\beta^T X$ is a convenient device to facilitate theoretical analysis, in practice sufficient dimension reduction still makes sense even without this assumption — it simply picks up those directions in X that are most important for predicting Y. In this respect it is very much like Principle Component Analysis (PCA), which ranks the importance of directions in a multivariate data set in explaining its variation. Sufficient dimension reduction simply ranks the linear combinations of predictors according to their ability to predict the responses.

2 Dimension reduction methods

A rigorous formulation of a dimension reduction problem is to find linear combinations $\beta^T X$ such that Y and X are independent conditioning on $\beta^T X$. In symbols

$$Y \perp\!\!\!\perp X|\beta^T X. \tag{2}$$

Here, only the column space of β matters, because the above statement is unaffected if we replace $\beta^T X$ by $(\beta A)^T X$ for any nonsingular $q \times q$ matrix A. The column space of β, then, is called a dimension reduction space. Under mild conditions, it can be shown the intersection of two dimension reduction spaces is again a dimension reduction space. The intersection of all dimension reduction subspaces is called the Central Space, and is written as $\mathcal{S}_{Y|X}$. Thus the goal of sufficient dimension reduction as inferring about $\mathcal{S}_{Y|X}$.

A property of dimension reduction subspaces that has practical importance is the following invariance. Let A be any $p \times p$ nonsingular matrix and c is a p-dimensional vector. Then

$$Y \perp\!\!\!\perp X|\beta^T X \Leftrightarrow Y \perp\!\!\!\perp X|\beta^T A^{-1}(AX + c) \Leftrightarrow Y \perp\!\!\!\perp X|(A^{-T}\beta)^T(AX + c),$$

where A^{-T} denotes $(A^{-1})^T$. This means if $\mathcal{S}$ is a dimension reduction space for $Y|X$, then $A^{-T}\mathcal{S}$ is a dimension reduction space for $Y|AX$. This allows us to perform dimension reduction on any affine transformation of original data.

A typical dimension reduction method consists of two parts: to estimate $\mathcal{S}_{Y|X}$ assuming its dimension q is known, and to estimate q. Classical dimension reduction methods include: Ordinary Least Square (OLS; Li and Duan (1989)), Principle Hessian Directions (PHD; Li (1992)), Sliced Inverse Regression (SIR; Li (1991), see also Zhu and Fang (1996)), and Sliced Average Variance Estimator (SAVE; Cook and Weisberg (1991)). They are based on the facts that, under some conditions, the column space of the following matrices

$$E(XY), \quad E(XX^TY), \quad \mathrm{cov}[E(X|Y)], \quad \mathrm{var}[I_p - \mathrm{var}(X|Y)] \tag{3}$$

are contained in $\mathcal{S}_{Y|X}$.

To estimate the central subspace we simply replace the expectations in the above quantities by their corresponding sample moments. A common strategy is to first transform the data into the standard scale, carry out the dimension reduction, and then transform the data back to the original scale. More specifically we follow these steps (again taking SIR as an example):

1. Standardize $X_1, \ldots, X_n$ to $\widehat{X}_1, \ldots, \widehat{X}_n$ so that the latter have sample mean 0 and sample variance I_p. That is, let μ_n and Σ_n be the sample mean and variance of X, and let $\widehat{X}_i = \Sigma_n^{-1/2}(X_i - \mu_n)$.

2. Let $\{I_1, \ldots, I_h\}$ be a partition of the range of Y, and $\widehat{Y}$ be the discretized Y along this partition. That is, $\widehat{Y}$ takes a distinct constant value on each interval I_j. Compute the SIR matrix

$$\widehat{K} = \mathrm{cov}_n[E_n(\widehat{X}|\widehat{Y})],$$

where $E_n(\cdot|\widehat{Y}=c)$ is the sample mean of $\widehat{X}$ on the set $\{\widehat{Y}=c\}$, and cov_n is the sample covariance of $E_n(\widehat{X}|\widehat{Y}=c_j)$, $j=1,\ldots h$. Let $\hat{v}_1,\ldots,\hat{v}_q$ be the eigenvectors of $\widehat{K}$ corresponding to its largest eigenvalues.

3. Transform back to the original scale by letting $\hat{w}_i=\Sigma_n^{-1/2}\hat{v}_i$. The span of these $\hat{w}_i$'s would then be taken as the estimate of $\mathcal{S}_{Y|X}$.

A commonly used procedure to determine q is sequential tests. For example, let K be the population SIR matrix (the third object in (3)), and consider the following hypotheses:

$$H_0:\mathrm{rank}(K)=k,\ k=0,1,2,\ldots$$

The dimension q is estimated by the first integer at which the above hypothesis is accepted. Denote the estimate of q by $\hat{q}$.

Once $\hat{w}_1,\ldots,\hat{w}_{\hat{q}}$ estimated, we use

$$X_i^*=(\hat{w}_1^T(X_i-\mu_n),\ldots,\hat{w}_{\hat{q}}^T(X_i-\mu_n))^T$$

as the dimensionally reduced predictor. The usual statistical analysis, such as diagnostics, regression, and classification, can be carried out using these new predictors. Based on our experience, $\hat{q}$ is usually within the range of 1 to 3. So we do have substantial saving for the subsequent analysis.

3 Recent developments

Recent years have witnessed a surge of research activities in this field, partly because of the demands for methods of processing high dimensional data, and partly because the richness of challenging problems in this field. In this section we briefly highlight of recent researches and references, as an assistance to researchers interested in this field.

3.1 Central Space and Central Mean Space

The conditional independence (2) is equivalent to $f_{Y|X}=f_{Y|\beta^TX}$, where f denotes conditional density functions. Thus we seek to reduce the dimension of X in the conditional distribution. In application, some aspects of the conditional distribution may be of the most interest, and should receive priority in estimation. The issue is analogous to parameter of interest versus nuisance parameter in classical inference.

It is possible to re-formulate dimension reduction to accommodate this distinction. For example, if the conditional mean $E(Y|X)$ is of primary interest in a

regression analysis, then instead of (2) we can ask if there is a $p\times q$ matrix β such that

$$Y \perp\!\!\!\perp E(Y|X)|\beta^TX.$$

Cook and Li (2002) studied this problem, and introduced Central Mean Space $\mathcal{S}_{E(Y|X)}$ as the smallest span(β) such that the above relation is satisfied. They studied theoretical properties of Central Mean Space and developed various means to estimate it. They established that OLS and PHD actually targets the Central Mean Space, whereas SIR and SAVE targets the Central Space itself.

A particularly interesting property is that if we let H be the Hessian matrix (which is the second object in (3)), then $\mathcal{S}_{E(Y|X)}$ is an invariant subspace of H. Based on this they introduced Iterative Hessian Transformation (IHT) method to estimate $\mathcal{S}_{E(Y|X)}$, which requires fewer assumptions than PHD and often outperforms OLS and PHD in estimating the Central Mean Space. Yin and Cook (2002) studied the more general problem where the parameter of interest is an arbitrary conditional moment. Cook and Li (2004) introduced sequential tests to determine the dimension the dimension of the Central Mean Space.

Since the characteristic function of $Y|X$ is also a form of conditional expectation, the Central Mean Space of $E(e^{it^TY}|X)$ can be used to estimate the Central Space. This is the approach used in Peng (2005), which introduced an effective method of estimating the Central Space.

3.2 Categorical predictors

Chiaromonte, Cook, Li (2002) and Li, Cook, Chiaromonte (2003) studied the dimension reduction problems where the predictor contains a categorical component. Here, we should first note that, because of the asymptotic normality of the marginal distributions, sufficient dimension reduction methods do apply directly to discrete (but numerical) predictors as long as p is reasonably large.

But when X is completely categorical, dimension reduction can be used to reduce the dimension of the continuous part of the predictor. The basic idea is this: Suppose the predictor is (X,W), where X is continuous but W is categorical, and consider the conditional independence:

$$Y \perp\!\!\!\perp X|(\beta^TX,W).$$

The smallest span(β) that satisfies this relation is called partial dimension reduction space (or Partial Space), and is denoted by $\mathcal{S}_{Y|X}^{(W)}$.

The estimation of the partial dimension reduction space is based on the following fact: let $\mathcal{S}_w$ be the dimension reduction space for each category, say $w=$

$1, \ldots, c$. Then

$$\mathcal{S}_{Y|X}^{(W)} = \mathcal{S}_1 + \cdots + \mathcal{S}_c.$$

Thus, we can apply the classical dimension reduction methods to each category, and incorporate them through sequential tests to estimate the Partial Space.

3.3 Efficiency and Exhaustiveness

Typically, a dimension reduction method, such as SIR, produces a set of preliminary vectors $\hat{\nu}_1, \ldots, \hat{\nu}_k$ that converge at $\sqrt{n}$-rate to $\nu_1, \ldots, \nu_k \in \mathcal{S}_{Y|X}$, where k is larger than the dimension of $\mathcal{S}_{Y|X}$. Traditional methods select from them the important vectors by Principal Component Analysis (PCA).

Recently, Cook and Ni (2005) introduced a more efficient way of putting these vectors together. The idea is to find a q-dimensional subspace (recall that q is the dimension of $\mathcal{S}_{Y|X}$) that is closest to these vectors. That is, to find a q-dimensional subspace $\mathcal{S}^*$ such that

$$\rho(\{\hat{\nu}_1, \ldots, \hat{\nu}_k\}, \mathcal{S})$$

is minimized, where $\rho(\cdot, \cdot)$ is a distance between a set of vectors and a subspace. The distance measure they used is

$$\min_{B,C} \left\{ \mathrm{vec}(\nu - BC)^T W \mathrm{vec}(\nu - BC) \right\},$$

where ν is the matrix $(\hat{\nu}_1, \ldots, \hat{\nu}_k)$, $B \in \mathbb{R}^{p \times q}$, and $C \in \mathbb{R}^{q \times k}$. Cook and Ni demonstrated substantial improvement in efficiency of this method over PCA.

Another challenging issue is exhaustive estimation. We have mentioned that under some conditions the four classical estimators converges to vectors in the Central Space (or Central Mean Space). But until recently there were no general results to guarantee that they are exhaustive — that is, the vectors to which they converge actually *span* the target spaces. In fact, it can be shown OLS, PHD, and SIR are not exhaustive. They perform well in detecting certain features in a regression function but are "blind" to some other features. Even when an estimator is exhaustive, it can be insensitive and inaccurate in estimating certain directions in the dimension reduction spaces. In this sense the issues of efficiency and exhaustiveness are connected. Several methods have been recently introduced partly to tackle this problem. For example, Hristache, Juditsky, Polzehl, Spokoiny (2001) and Xia, Tong, Li, Zhu (2002), and Li, Zha, and Chiaromonte (2005).

3.4 Multiple responses

Another area of recent advances is the extension of dimension reduction methods to the situations where the responses are also multivariate. This actually involves

two issues: how to reduce the dimension of X when Y is a random vector; how to reduce the dimension of Y itself.

Cook and Setodji (2003) demonstrated that, to estimate the Central Mean Space, one can find the dimension reduction space for each component of Y, and then combine them to estimate the dimension reduction space for Y versus X. This is based on the following fact: if $\mathcal{S}_j$ is the Central Mean Space for $Y_j|X$, $j = 1, \ldots, r$, then

$$\mathcal{S}_{E(Y|X)} = \mathcal{S}_1 + \cdots + \mathcal{S}_r.$$

Thus, we can use classical methods for single-response problem to estimate $\mathcal{S}_1, \ldots, \mathcal{S}_q$, and then combine them using sequential test. See also Yin and Bura (2005).

Regarding the dimension reduction for Y, Li, Aragon, Shedden and Agnan (2003) proposed to seek a linear combination of Y that can be best predicted by X. That is, we sequentially choose a vector θ and a function g to minimize the sample estimate of the following criterion:

$$\frac{E[\theta^T Y - g(X)]^2}{\mathrm{var}(\theta^T Y)},$$

where g ranges over all square-integral functions of X. Here, we can also first reduce the dimension of X as it appears in the classical context of (2).

There are many other important issues but it is impossible to cover them within this short note, for example: dimension reduction for variables contaminated with measurement errors (Carroll and Li, 1992; Lue, 2004); Using dimension reduction for model selection (Cook, 2004; Li, Cook, Nachtsheim, 2005); Other means of order determination than sequential tests, such as the bootstrap method (Ye and Weiss, 2003) and the BIC (Zhu, Miao, Peng, 2005), extension of dimension reduction methods to functional data sets (Ferre and Yao, 2003); as well as issues concerning outliers and robustness Cook and Critchley (2000), Gather, Hilker, and Becker (2002).

References

Carroll, R. J. and Li, K. C. (1992). Measurement error regression with unknown link: Dimension reduction and data visualization. *Journal of the American Statistical Association*, 87, 1040–1050.

Cook, R.D. (1994). On the interpretation of regression plots. *Journal of the American Statistical Association* **89**, 177–189.

Cook, R.D. (1996). Graphics for regressions with a binary response. *Journal of the American Statistical Association* **91**, 983–992.

Chiaromonte, F., Cook, R. D. and Li, B. (2002). Partial dimension reduction with categorical predictors. *The Annals of Statistics.* **30**, 475-497.

Cook, R.D. (1998). *Regression Graphics.* Wiley, New York.

Cook, R. D. and Critchley, F. (2000), Identifying regression outliers and mixtures graphically. *Journal of the American Statistical Association.* **95**, 781-794.

Cook, R.D. and Li, B. (2002). Dimension reduction for the conditional mean. *The Annals of Statistics* **30**, 455–474.

Cook, R.D. and Li, B. (2004). Determining the dimension of Iterative Hessian Transformation. *The Annals of Statistics* **32**, 2501–2531.

Cook, R. D. and Nachtsheim, C. J. (1994). Reweighting to achieve elliptically contoured covariates in regression. *Journal of the American Statistical Association* **89**, 592–599.

Cook, R.D. and Ni, L. (2005). Sufficient dimension reduction via inverse regression: A minimum discrepancy approach. *Journal of the American Statistical Association* **100**, 410–428.

Cook, R.D. and Setodji, C.M. (2003). A model-free test for reduced rank in multivariate regression. *Journal of the American Statistical Association* **98**, 340–351.

Diaconis, P. and Freedman, D. (1984). Asymptotics of graphical projection pursuit. *The Annals of Statistics.* **12**, 793–815.

Ferre, L. and Yao, A.F. (2003). Functional sliced inverse regression analysis. *Statistics* **37**, 475–488.

Gather, U., Hilker, T. and Becker, C. (2002). A note on outlier sensitivity of sliced inverse regression. *Statistics* **36**, 271-281.

Hall, P. and Li, K. C. (1993). On almost linearity of low dimensional projections from high dimensional data. *The Annals of Statistics,* **21**, 867–889.

Li, B., Cook, R. D. and Chiaromonte, F. (2003). Dimension reduction for conditional mean in regression with categorical predictors. *The Annals of Statistics.* **31**, 1636–1668.

Li, B., Zha, H., and Chiaromonte, F. (2005). Contour regression: a general approach to dimension reduction. *The Annals of Statistics.* **33**, 1580–1616.

Li, K. C. (1991). Sliced inverse regression for dimension reduction (with discussion). *Journal of the American Statistical Association* **86**, 316–342.

Li, K.C. (1992). On principal Hessian directions for data visualization and dimension reduction: Another application of Stein's lemma. *Journal of the American Statistical Association* **87**, 1025–1039.

Li, K. C. and Duan, N. (1989). Regression analysis under link violation. *The Annals of Statistics* **17**, 1009–1052.

Li, K. C., Aragon, Y., Shedden, K., Thomas-Agnan, C. (2003). Dimension Reduction for Multivariate Response Data. *Journal of the American Statistical Association* **98**, Vol. 98, 99-109.

Li, K.C. and Duan, N. (1989). Regression analysis under link violation. *The Annals of Statistics* **17**, 1009–1052.

Li, L., Cook, R.D., Nachtsheim, C. J. (2005). Model-free variable selection. *Journal of Royal Statistical Society, Series B* **67**, No.2, 285-299.

Lue H.-H. (2004). Principal Hessian Directions for regression with measurement error. *Biometrika*, 91. 409–423.

Xia, Y., Tong, H., Li, W. K., and Zhu, L. (2002). An adaptive estimation of dimension reduction space. *Journal of the Royal Statistical Society, Series B* **64**, 363-388.

Ye, Z. and Weiss R. (2003). Using the bootstrap to select one of a new class of dimension reduction methods. *Journal of the American Statistical Association* **98**, 968–979.

Yin, X. and Bura, E. (2005). Dimension reduction for multivaraite response in regression. *Journal of Statistical Planning and Inference*, to appear.

Yu Zhu and Peng Zeng (2006). Fourier Methods for Estimating the Central Subspace and the Central Mean Subspace in Regression. To appear in *Journal of the American Statistical Association.*

Zhu, L., Miao, B., Peng, H. (2006). On Sliced Inverse Regression with high-dimensional covariates. *Journal of the American Statistical Association* **101**, 630–643.

Zhu, L. and Fang, K.-T. (1996). Asymptotics for kernel estimate of sliced inverse regression. *Annals of Statistics* **24**, 1053–1068.

Singular Value Decompositions on Microarray Data

Jianhua Hu and Xuming He

Department of Biostatistics and Applied Mathematics, University of Texas M.D. Anderson Cancer Center (jhu@mdanderson.org)

Department of Statistics, University of Illinois at Urbana-Champaign (x-he@uiuc.edu)

In algebra, singular value decomposition (SVD) is an important factorization of a rectangular matrix, with several applications in signal processing and statistics. In statistics, the SVD has been playing an increasing role in high dimension data analysis, especially in microarray data analysis. In this article, we shall examplify the value of SVD in microarray data analysis.

Introduction to SVD

Let $\mathbf{Y}$ be a real m-by-n matrix of positive rank. Its singular value decomposition is

$$\mathbf{Y} = \mathbf{U}\,\mathbf{\Sigma}\,\mathbf{V}^{\mathrm{T}} \qquad (1)$$

where $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices (assumed orthonormal for identifiability), $\mathbf{\Sigma}=\mathrm{diag}(\sigma_1,\ldots,\sigma_r)$, $r=\min(m,n)$, and $\sigma_1 \geq \ldots \geq \sigma_r \geq 0$. The LAPACK Users' Guide (Anderson et al., 1999) describes the numerical routines pertaining to the SVD, from which we borrowed much of the notation. Boldface represents matrices and vectors of the corresponding scalar quantities. The σ_w are called the singular values, the square roots of r (usually) non-zero eigenvalues of $\mathbf{YY}^{\mathrm{T}}$ and $\mathbf{Y}^{\mathrm{T}}\mathbf{Y}$. The first r columns of $\mathbf{V}$ are the right singular vectors and the first r columns of $\mathbf{U}$ are the left singular vectors, ordered according to the descending singular values. The column vectors of $\mathbf{\Sigma}\,\mathbf{V}^{\mathrm{T}}$ are called the characteristic modes of $\mathbf{Y}$. Very often, the major characteristics of the data matrix can be understood through the first few characteristic modes.

Introduction to Microarray Data

The high-density oligonucleotide and cDNA microarray technology has been widely used in many areas of biomedical research. Gene expression profiling or microarray analysis has enabled the measurement of thousands of genes in a single RNA sample. The microarray data are characterized by high dimensions, which makes it necessary to focus on the first few structures. Here, a data matrix may refer to gene expression indices where the rows correspond to genes and the columns correspond to arrays, or intensity measures of a given probe-set with rows for probes and columns for arrays.

SVD for Gene Expression Index

The Li-Wong model is commonly used in estimating gene expression indices from the Affymetrix data. For example, we consider the reduced mode for the difference **PM** – **MM**

$$PM_{ij} - MM_{ij} = \theta_i\phi_j + \varepsilon_{ij}. \qquad (2)$$

where PM_{ij} is the perfect match intensity and MM_{ij} is the mis-match intensity of the i-th array and j-th probe. A constraint $\Sigma_j\phi_j^2=J$ is sufficient to provide identifiability of the gene expression indices θ_i . There is a direct relationship between the first characteristic mode of the SVD and the above model. As shown in Hu et al. (2006), the gene expression index estimates from this model is simply proportional to the first characteristic mode of $\mathbf{Y}=\mathbf{Y}_{PM}-\mathbf{Y}_{MM}$, the data matrix that consists of PM – MM . One advantage of using the SVD equivalent is that faster SVD algorithms (Anderson et al. 1999) are considerably faster than the iterative likelihood method employed in dChip.

In simpler languages, we see that the Li-Wong model captures only the first singular structure of the array-by-probe data matrix. If a uni-dimensional summary will be used for at least the majority of the genes, we can ask how the expression index estimates can be improved. One solution is to maximize the nontrivial information contained in the first singular structure. Towards this goal, we may use transformations on **Y**, or consider a two-dimensional summary for gene expression.

Examining the patterns in the first singular vectors in both the left and right singular matrices can also help identify probe irregularities. This might be particularly useful in analyzing the exon tiling arrays, but work remains to be done.

SVD for Gene Expression Patterns

Holter et al. (2000) analyze several different types of gene expression data by using the SVD with the genes representing the rows of **Y** and the arrays over different time points of biological activity representing the columns. They find that the characteristic modes largely reflect the genome-wide expression pattern and are not gene-specific. For a specific gene, the temporal pattern of variation in expression can be expressed as a linear combination of the characteristic modes, with gene-specific coefficients specified in **U**. The contribution of each mode to the final gene expression profile progressively diminishes from the lower to the higher order modes for the gene expression data sets, but should be approximately equal for a random data set. They demonstrate that the major features of the overall genetic response of the cells are contained in a combination of just a few different modes. In their data, the first two modes dominate the genome-wide expression pattern, with a very simple structure applicable to most genes. The remaining modes describe minor elements in the patterns, with a considerable fraction due to small scale fluctuations and experimental noise. They also point out that the first two dominant modes are robust in the sense that their shapes do not change substantially upon removal of the last three time points in their experiments.

Extending their earlier ideas, Holter et al. (2001) propose a time translational matrix by modeling expression within a linear framework, using the characteristic modes obtained from SVD. By applying the time translational matrix on the time evolution gene expression data, the future expression levels of genes can be predicted based on the expression at some initial time. The resulting time translational matrix provides a measure of the relations among the characteristic modes and governs the time evolution. They show that a truncated matrix involving only a few modes is a good approximation of the full time translation matrix. Again it suggests that the number of essential patterns among the genes is small.

There are some other applications of SVD in microarray experiments. For example, Alter et al. (2000) adopt SVD for processing and modeling expression data . Ghosh (2002) uses SVD to develop predictive models for correlating gene expression data with clinical outcomes. Liu et al. (2002) propose a clustering method with robust SVD and segmentation.

SVD for Normalization

The high-density oligonucleotide and cDNA microarray technology has been widely used in many areas of biomedical research. A variety of technical sources including differences in sample concentrations and hybridization conditions can result in variations from one array to another, and yet these variations are biologically uninteresting. To make comparable the gene expression intensities of all the arrays, a normalization procedure is routinely used prior to any statistical analysis of the gene expression data. While the necessity of normalization is almost universally agreed upon in the scientific community, questions remain about the quality and effectiveness of normalization procedures; see Lyons-Weiler (2003). Research on normalization has been focusing on what can be gained, without asking what has been missed.

If we view a normalization procedure as decomposing the non-normalized data $\boldsymbol{Y}$ into *normalized data* $\boldsymbol{N}$ plus *residuals* $\boldsymbol{R}$, that is, $\boldsymbol{Y}=\boldsymbol{N}+\boldsymbol{R}$, then SVD can be used on both **N** and **R** . By taking only the first few eigen-structures of **N** that are important, we can reduce noise. By asking if the first few eigen-structures of R still contain useful gene expression patterns, we may recover what is lost in a given normalization procedure. Based on this idea, Hu and He (2006) proposed an enhancement to the quantile normalization with the objective of minimizing information loss.

We consider one example to demonstrate the value of the enhancement. The data set was from an experiment conducted by the Division of Human Cancer Genetics at the Ohio State University (Lemon et al., 2002). It was a set of 18 HuGeneFL arrays, each of which was loaded with 11 ug/200uL labeled cRNA. The Affymetrix GeneChip HuGeneFL Array is a single array that enables the relative monitoring of mRNA transcripts of approximately 5,600 full-length human genes. There were 7129 probe sets in each array. From the human fibroblast cell experiment, a set of *stimulated* and *starved* samples was produced. Another RNA sample was produced as a balanced mixture of simulated and starved samples, which was called the *50:50 mixture*. For each condition (serum stimulated, starved and a 50:50 mixture), two aliquots of RNA were drawn and processed separately on three consequent days. In addition, spiked-in genes were added in the following way: *Lys* and *Phe* RNAs at 0.08 ng/8 μg were added to the stimulated samples, the same amount of *Dap* and *Thr* were added to the starved sample, and these four RNAs at 0.04 ng/8 μg each were assigned to the 50:50 mixture. Another set of control genes, *BioB*, *BioC*, *BioD* and *Cre* with final concentrations of 1.5, 5, 25 and 100 pM, respectively, were also added to all samples. At each condition, six replicated HuGeneFL arrays were produced. Minimization of the technical variability was realized using a single fluidics station and a same lot for the 18 arrays.

The fundamental expression pattern across the arrays is known in this study. Arrays 1-6 used the 50:50 mixture sample, arrays 7-12 used the stimulated samples, and arrays 13-18 used the starved samples. Since we know in advance the true concentration levels of the mRNAs of the spiked-in control genes in this experiment, we can assess the relationship between the underlying amount of total RNA in each sample with the estimated expression indices of those genes. Among the spiked-in genes for example, *Dap* and *Thr* obtained 0, 0.04, and 0.08 ng/8μg total RNA in the stimulated, mixture, and starved samples, respectively. We plotted the estimated gene indices for two probe sets in Figures 1-4, one for *Dap* and the

other for *Thr*. In both cases, the quantile normalization method could not differentiate the expression levels between the 50:50 mixture and the starved samples, while the enhanced method succeeded in doing so. This phenomenon persisted for all of the 12 spiked-in genes.

The quantile normalization method failed to differentiate the expression levels between the 50:50 mixture and the starved samples, because the pre-normalization expression indices for those probes were about the highest in arrays 1-6 as well as in arrays 13-18. The post quantile normalization expression indices were simply the average over those arrays. The proposed enhancement avoided this undesirable property by allowing the array to array profile patterns to persist even for those genes with the highest expression genes. This was achieved by de-noising as well as recovering the first singular structure from the RMA residual matrix.

Conclusions

Microarray data are unarguably high dimensional. In any typical study, the number of genes is measured in thousands. Even for a single gene, one needs to analyze multiple arrays and multiple probe-level data to arrive at any conclusions about the gene expression profile. Low sample size and high dimension are a very challenging dual for statistical analysis of the data. Singular value decomposition and other forms of low-rank approximation to the data matrices are powerful tools to address the challenge.

The SVD is a powerful mathematical tool, yet it does not solve all the problems in microarray data analysis. The SVD itself does not lead to biological interpretations for the major singular vectors. New statistical methods are needed, for example, to test the adequacy of a uni-dimensional structure of the data matrix without relying on the usual asymptotics in multivariate statistics. It might also be important to understand the biological implications of what is beyond the first singular structure in a microarray data matrix. The potential of SVD may be best realized as statisticians and the biological scientists continue to work together to embrace new technology and innovative methodology.

Reference

Alter, O., Brown, P. O. and Botstein, D. (2000). Singular value decomposition for genome-wide expression Data processing and modeling. *Proceedings of the National Academy of Science USA*, 97, 10101-10106.

Anderson, E., Bai, Z., Bischof, C., Blackford, S., Demmel, J., Dongarra, J., Du Croz, J., Greenbaum, A., Hammarling, S., McKenney, A. and Sorensen, D. (1999). *LAPACK Users' Guide* (3rd ed.): the Society for Industrial and Applied Mathematics.

Ghosh, D. (2002). Singular value decompostion regression models for classification of tumors from microarray experiments. *Proceedings of the 2002 Pacific Symposium on Biocomputing*, 18-29.

Holter, N. S., Mitra, M., Maritan, A., Cieplak, M., Banavar, J. R. and Fedoroff, N. V. (2000). Fundamental patterns underlying gene expression profiles: simplicity from complexity. *Proceedings of the National Academy of Science USA*, 97, 8409-8414.

Holter, N. S., Maritan, A., Cieplak, M., Fedoroff, N. V. and Banavar, J. R. (2001). Dynamic modeling of gene expression data. *Proceedings of the National Academy of Science USA*, 98, 1693-1698.

Hu, J. and He, X. (2006). Enhanced quantile normalization of microarray data to reduce loss of information in the gene expression profilet. To appear in *Biometrics*.

Hu, J., Wright, F. A. and Zou, F. (2006). Estimation of expression indexes for oligonucleotide arrays using the singular value decomposition. *Journal of the American Statistical Association*, 101, 41-50.

Lemon, W. J., Palatini, J. J. T., Krahe, R. and Wright, F. A. (2002). Theoretical and experimental comparisons of gene expression indexes for oligonucleotide arrays. *Bioinformatics*, 18, 1470-1476.

Liu, L., Hawkins, D. M., Ghosh, S. and Young, S. S. (2003). Robust singular value decomposition analysis of microarray data. *Proceedings of the National Academy of Science USA*, 100, 13167-13172.

Lyons-Weiler, J. (2003). Profound normalisation challenges remain in the analysis of data from microarray experiments. Applied Bioinformatics, 2, 193-195.

Dr. Janhua Hu is Assistant Professor of Bioinformatics and Biostatistics at the University of Texas M. D. Anderson Cancer Center. Dr. Xuming He is Professor of Statistics, University of Illinois at Urbana-Champaign, and is currently Visiting Professor at the Department of Biostatistics and Applied Mathematics, University of Texas M. D. Anderson Cancer Center.

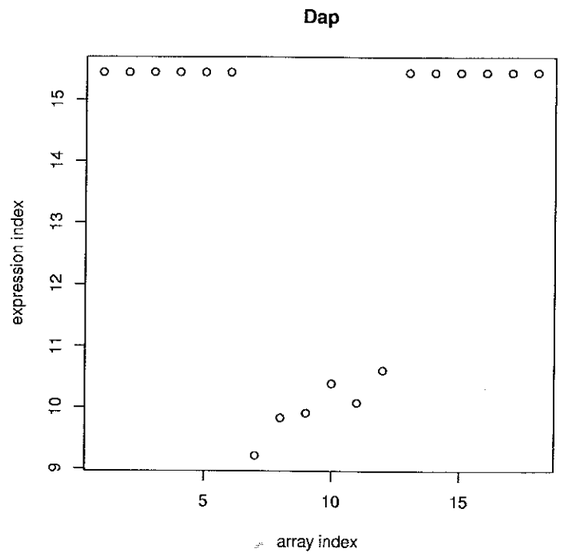


Figure 1: Human fibroblast data: Scatter plot of expression indices of *Dap* obtained through the quantile normalization. The array index, 1 to 6, 7 to 12, and 13 to 18, correspond to the 50:50 mixture, stimulated and starved samples.

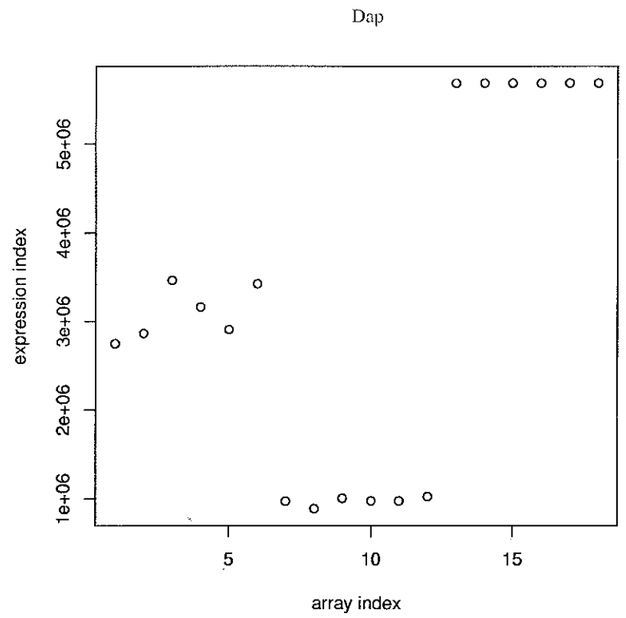


Figure 2: Human fibroblast data: Scatter plot of expression indices of *Dap* obtained through the enhanced normalization. The array index, 1 to 6, 7 to 12, and 13 to 18, correspond to the 50:50 mixture, stimulated and starved samples.

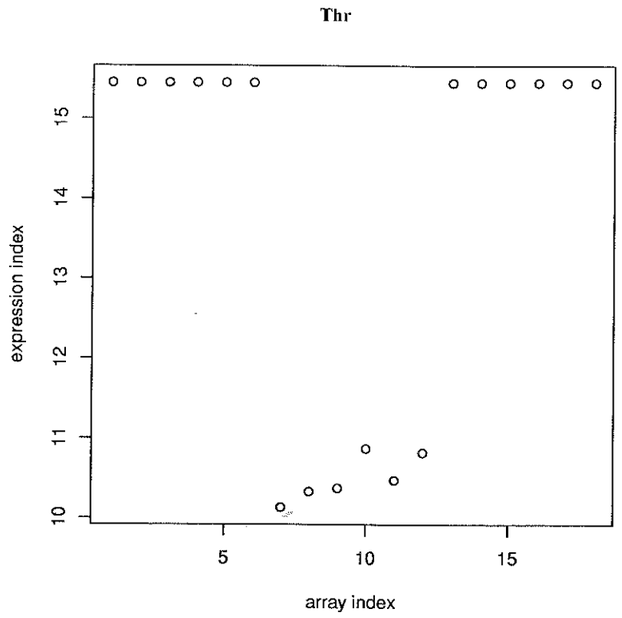


Figure 3: Human fibroblast data: Scatter plot of expression indices of *Thr* obtained through the quantile normalization. The array index, 1 to 6, 7 to 12, and 13 to 18, correspond to the 50:50 mixture, stimulated and starved samples.

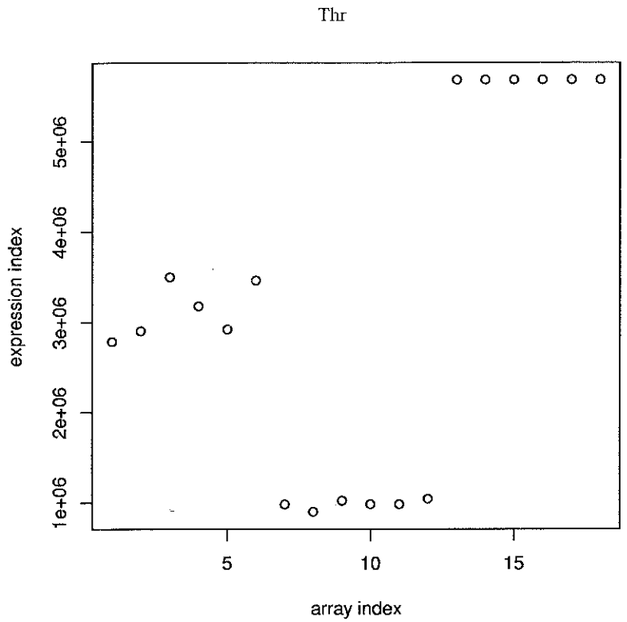


Figure 4: Human fibroblast data: Scatter plot of expression indices of *Thr* obtained through the enhanced normalization. The array index, 1 to 6, 7 to 12, and 13 to 18, correspond to the 50:50 mixture, stimulated and starved samples.

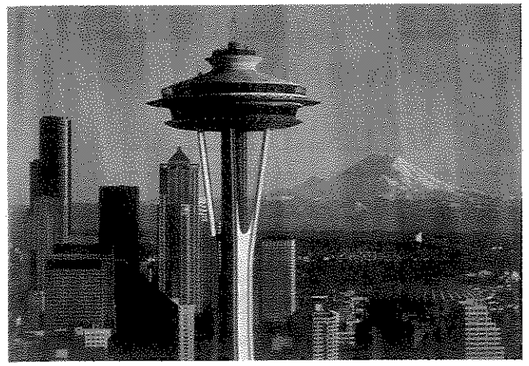
2006 ICSA Annual Banquet – Seattle, Washington

On behalf of the organizing committee for the ICSA annual banquet, I would like to welcome you to Seattle! Following tradition, ICSA will have its annual banquet after the ICSA Annual Members Meeting during JSM on Wednesday, August 9. Our annual banquet will be held at the New Kowloon Seafood Restaurant in Seattle's Chinatown / International District. It is easily accessible by a number of buses from downtown. Ms. Li Ma, an internationally renowned Guzheng virtuoso and winner of the Award of Excellent Performance in the 1995 Oriental Cup of National Juvenile Guzheng, will perform during dinner. Tickets can be purchased at ICSA booth in JSM and the cost of $20.00/person. Information about the restaurant can be found at http://new-kowloon.cwok.com/new-kowloon/index_e.html.

Seattle has many local attractions within walking distance of downtown. Watch fish fly through the air at Pike Place Market, view the city and Mt. Rainer from the Space Needle, release your inner rock star at the Experience Music Project (EMP), experience the wonders of the 400,000 gallon Seattle Aquarium on the waterfront, or tour the Seattle harbor on a cruise departing from the waterfront. Other attractions outside of downtown include watching the Mariners from Safeco Field, enjoying a traditional Northwest Coast Indian Style salmon dinner at Tillicum Village on Blake Island (Argosy Boat Tours), or traveling to Mt. Rainer (1.5/hour drive). Grey Line offers a variety of bus tours in and around Seattle for those who don't want to drive http://www.graylineofseattle.com/.

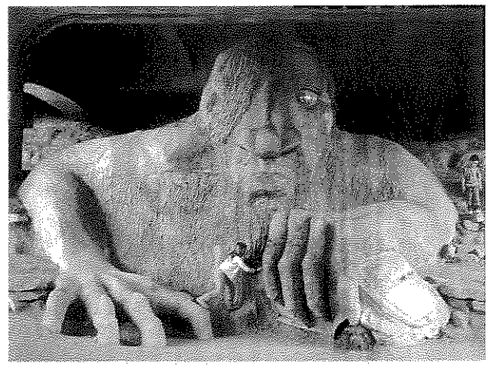
Pictures about Seattle are shown next page.

Xiao-Hua Andrew Zhou
Chair, Organizing Committee
Professor, Department of Biostatistics, School of Public Health
Adjunct Professor, Department of Psychiatry and Behavior Sciences
School of Medicine
Director and Investigator
Biostatistics Unit, HSR&D Center of Excellence
VA Puget Sound Health Care System
University of Washington
Office F644, HSB Box #357232, Seattle, WA 98198
Phone: 206-277-3588, e-mail: azhou@u.washington.edu



Seattle, Washington

Seattle's Famous Pike Place Market



"Troll Under the Bridge"—Fremont Neighborhood

ICSA 2007 Applied Statistics Symposium

June 3 - 6, 2007 at Raleigh, NC, U.S.A

The ICSA Applied Statistics Symposium Program Committee invites you to participate the 16th annual ICSA Applied Statistics Symposium at Research Triangle Area in North Carolina.

DATE: June 3 to 6, 2007 (Sunday to Wednesday).

PROGRAMS: Short courses on June 3 and technical sessions on June 4 to June 6. The program tentatively consists of 5 short courses (Mixed Effects Models, Adaptive Designs, Genomics, Survival Analysis, and SAS) and about 40 invited sessions.

LOCATION AND ACOMMODATION: **Sheraton Capital Center Hotel** in downtown Raleigh, North Carolina.

KEYNOTE SPEAKERS: Dr. **Janet Woodcock** (FDA Deputy Commissioner) and Professor **LJ Wei** (Harvard University).

CALL FOR PAPERS: The program committee invites talks on all aspects of statistics. Abstracts for the contributed papers are due **March 31, 2007**. Please submit abstracts via e-mail to: Professor Danyu Lin, University of North Carolina at Chapel Hill (lin@bios.unc.edu). The abstract should include name, affiliation, mailing address, telephone number, and e-mail address of the author(s), and not exceed 200 words.

ICSA STUDENT AWARDS AND TRAVEL GRANTS: The Symposium will sponsor up to four student awards and travel grants, including the **J.P. HSU MEMORIAL STUDENT AWARD**. The submission deadline is **March 15, 2007**, please read details on the separate page in this issue.

2007 APPLIED STATISTICS SYMPOSIUM STEERING COMMITTEE:
Shuyen Ho (co-chair, shu-yen.ho@gsk.com), Danyu Lin (co-chair, lin@bios.unc.edu), Shein-Chung Chow, Qiming Liao, Patrick Liu, Yu Lou, Sue-Jane Wang, Linda Yau, Zhao-Bang Zeng, Henry Zhao, Hui Zhi, HaiBo Zhou

ICSA 2007 Applied Statistics Symposium
Student Awards & Travel Grants

The 2007 Annual ICSA Applied Statistics Symposium will be held during **June 3-6, 2007** at the **Sheraton Capital Center Hotel in downtown Raleigh, North Carolina, USA**. The Symposium will sponsor the Student Awards and Travel Grants. The main purpose of the award is to encourage student members of ICSA to participate and present their research work at this annual meeting.

Qualifications: The student must be an ICSA member (or join at the time of manuscript submission), a degree candidate in any term during the academic year 2006-2007 at an accredited institute and be able to register and present the work at the 2007 symposium.

Manuscripts should be prepared double spaced using Biometrics or JASA guidelines for authors. They must be no more than 20 pages in length exclusive of tables and figures. Use one-inch margins and no smaller than 12 point type. The work must be relevant to applications in a variety of fields including biomedicine, business, etc. The manuscript may be co authored with a faculty adviser and/or a small number of collaborators, although the student must be the first author.

Review and Selection Process: The members of Student Award Committee and J. P. Hsu Memorial Scholarship Committee will receive blinded copies of the submitted manuscripts from the Committee Chairs and review them based on the following criteria:

- The manuscript should be well motivated by an application relevant to the specific field(s).
- The methodology developed should be applicable to the motivating problem. Inclusion of an application of the proposed methodology to a particular study will be favorably considered.
- Organization and clarity of the presentation will be considered as well.

Up to 4 award winners will be selected by the Awards Committees.

Each winner will receive a certificate, $400, and tuition for one short course of his/her choice. Winners will be notified around **April 15, 2007**.

Submission of Manuscripts: Manuscripts should be received and postmarked no later than **March 15, 2007.** The submission should include:

- A cover letter
- One complete title page with author(s), institutional affiliation, mailing address, phone/fax numbers and email address
- Five copies of the manuscripts with only a title, but no information on authors or affiliation, on the first page
- Two copies of abstract
- Two copies of the ICSA membership application for non-members

Membership forms can be downloaded from http://www.icsa.org. All materials should be mailed to:

Professor Shein-Chung Chow (sheinchung.chow@duke.edu)
Department of Biostatistics and Bioinformatics
2400 Pratt Street
Room 0311 Terrace Level
Duke University
Durham, NC 27705
USA

International Chinese Statistical Association

Profit and Loss

January 1, 2006 through June 30, 2006

Balance, Dec 2005	**61513.92**
Income	
Membership Dues	3840.00
Total Income	**3840.00**
Expense	
Miscellaneous	
Starting fund for dinner and member meeting at JSM 2006	800.00
Miscellaneous[1]	150.00
Total Miscellaneous	950.00
Postage and Delivery	
January Bulletin	1613.76
Total Postage and Delivery	1613.76
Printing and Reproduction	
January Bulletin	3817.50
Total Printing and Reproduction	3817.50
Web Page Hosting	107.40
Total Expense	**6488.66**
Net Ordinary Income	-2648.66
Other Income/Expense	0
Net Other Income	0
Net Income	-2648.66

International Chinese Statistical Association

Balance Sheet

January 1, 2006 through June 30, 2006

ASSETS	
Checking/Savings	
Checking	17978.46
Savings-Money Market CD[2]	40886.80
TOTAL ASSETS	**58865.26**
LIABILITIES & EQUITY	
Equity	
Opening Balance Jan 1, 2006	61513.92
Net Income	-2648.66
TOTAL LIABILITIES & EQUITY	**58865.26**

Note:

1. Check cleared for postage of book/journal donation in 2005.
2. 2006 interest income from CD not included.

Submission Guidelines for ICSA Bulletin

Articles

The International Chinese Statistical Association (ICSA) Bulletin welcomes the submission of articles by our members. Articles submitted should be written in English. The contents should be aimed at the general reading.

Articles should be submitted electronically in Microsoft Word documents. Files with Times New Roman font size of 12 points are preferred. Leave ¾ inches of blank space on each margin in regular or letter size pages (8.5 inches in width by 11 inches in height). If photo pictures are included in the articles, the corresponding JEPG files must be attached.

Submission deadlines are December 15 for the January issue, and June 15 for the July issue. Articles received after the deadline will be published in the following issue of Bulletin.

Reports or Announcements

The submission guidelines of reports or announcements are the same as the articles. Editable files in Microsoft Word are preferred. If needed, EXCEL may be used. If PDF files are sent, be sure that the margins or fonts are met as required in the Articles.

Advertisements

The ICSA Bulletin welcomes the submission of advertisements related to the profession of Statistics. The fee charge is $225 for a half page, $325 for a full page, $350 for a color page (in front-inside, bottom-inside, or bottom-outside pages). If it is a color page, the corresponding PDF file for printing use must be sent. If you have questions or advertisement opportunities, please contact the Advertising Manager or Editor-in Chief.

Questions

Please submit your questions to the Editor-in-Chief by email: tkao@usuhs.edu

International Chinese Statistical Association
泛華統計協會

Membership Application & Renewal Form

Name	(Last)	(Middle)	(First)
(English)			
(Chinese)			
Address			
Office	Address:		
	City:		
	State:	Zip Code:	Country:
	Email:	Telephone:	FAX:
Home	Address:		
	City:		
	State:	Zip Code:	Country:
	Email:	Telephone:	FAX:
Education			
	Degree:	Year Graduated:	
	University:		
Professional Occupation & Title			
	Occupation:	Title:	

Membership Fees

Regular	(US$40)
Student	(US$20)
Permanent	(US$400)
Spouse	(50%)
Donations	
Total Amount Paid:	US$

Statistical Area of Interest (circle all applicable):

A: Agriculture	B: Business / Economics
C: Computing / Graphics	D: Education
E: Engineering	F: Health Sciences
G: Probability	H: Social Sciences
I: Biostatistics	N: Theory & Methodology

Please Make Check Payable to: I.C.S.A. Mail This Form & Fees to:
ICSA c/o Ivan S. F. Chan, 6 Sarah Court, Dresher, PA 19025