International Chinese Statistical Association

http://www.icsa.org

泛華統計協會

Features:

Interview with a Distinguished Statistician

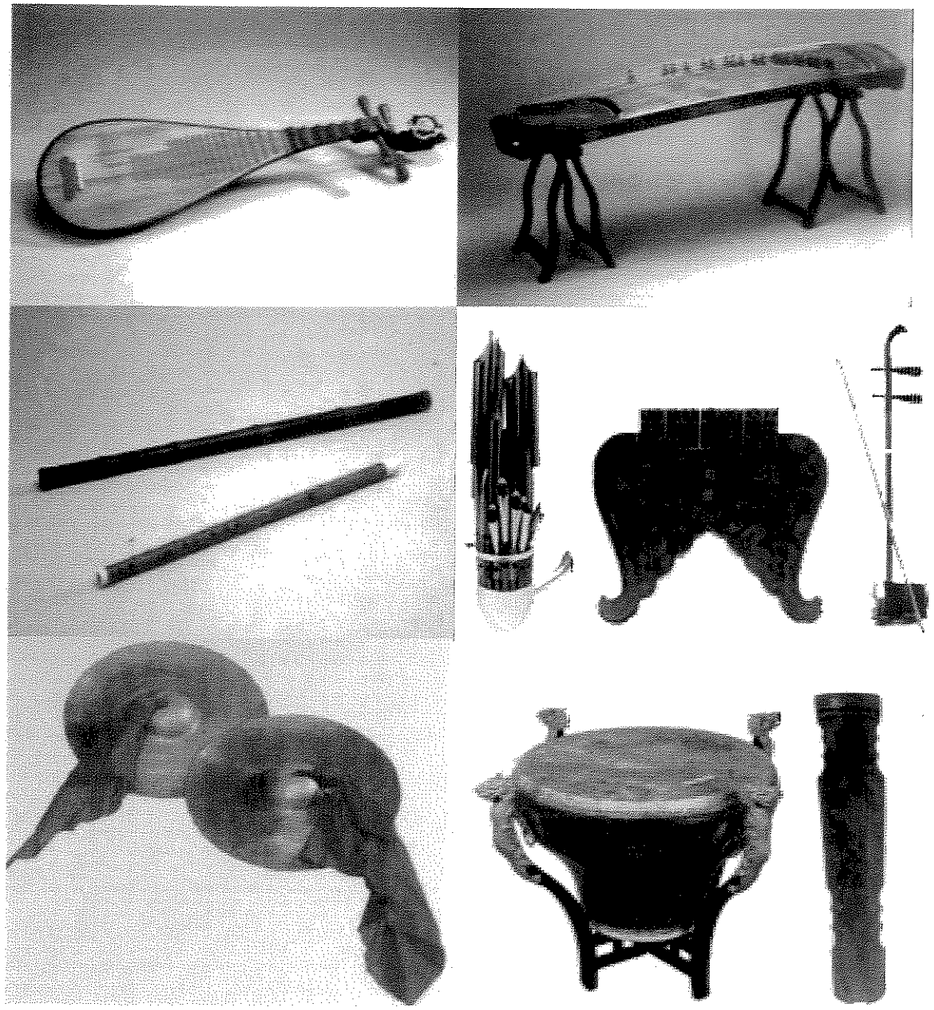
Candidates of ICSA Officers

Statistical Issues

Meeting Announcements

Highlights of 2004 Applied Statistics Symposium

Bulletin July 2004



Chinese Music

Music is one of the four arts in ancient China to symbolize the classic scholarship. The history of Chinese music can be dated back to about 7000 years ago. Chinese music instruments may be classified into string, wind, and percussion instruments. They were used in all kind of occasions to describe the civilian life events and even the national warfare through out the Chinese history.

From the Editor

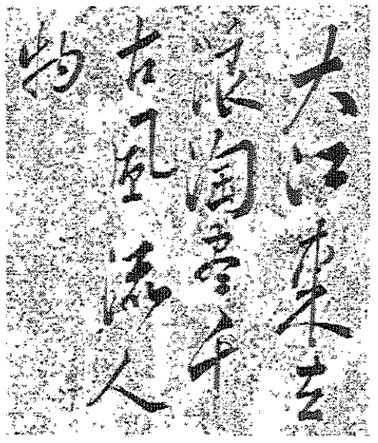
Kao-Tai Tsai, Ph.D.

As in the ancient Chinese lyric poems, "The river rolls toward the East and washes away the heroes with its waves and bubbles." This year, the ICSA lost a wonderful friend and a strong supporter, Dr. Jiann-Ping Hsu. During her long career as a statistician and entrepreneur, she set the high moral standard in various aspects of human life. Her beautiful journey, her enormous intellectual capacity, and her genuine sincerity deeply impressed the people she had ever touched and will be forever remembered and celebrated through the endowment established by her beloved husband, Dr. Peace.

The curtain of the ICSA 2004 Applied Statistics Symposium had finally closed after last year's interruption of SARS. The record attendance and the rich program indicated the growth of the ICSA beyond a local phenomenon. This was the second time the Symposium had ever taken place outside the East coast of the US. Incidentally, the first time this took place in Chicago, it was co-chaired by Dr. Jiann-Ping Hsu. We congratulate the success of the Symposium Committee and thank them for the great effort.

In this issue, we are extremely fortunate to have the opportunity to interview Dr. Kettenring, a past President of the ASA, to share his experience of being the President of the largest statistical association and his insight of a successful statistical career.

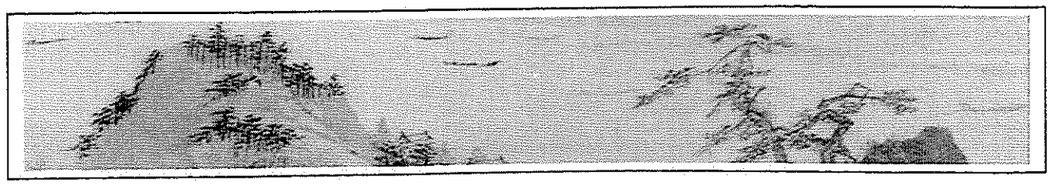
We would also like to congratulate Professor C. F. Jeff Wu of being elected to the National Academy of Engineering for "conceiving and building modern systems of experimental design based on contemporary methods for parameter estimation to provide quality improvements." This is indeed a well-deserved honor for his life long professional dedication.



We hope you enjoy the fruit of our labor in this issue. And, as usual, your help in any ways will be greatly appreciated.

Table of Contents
ICSA Bulletin, July 2004

Editorial Members
ICSA Bulletin, July 2004



Committee Chairs
ICSA, 2004

Board of Directors

Ivan Chan	Frank Shen
Jiahua Chen	Naitee Ting
Ming-Hui Chen	Sue-Jane Wang
Rong Chen	Wayne Wang
Jiqian Fang	Greg Wei
Xumin He	Qiwei Yao
Shu-Yen Ho	Zhiliang Ying
Zhaohai Li	Heping Zhang
Xihong Lin	Hongyu Zhao

Standing Committee
Program Committee
Naitee Ting
Finance Committee
Weiying Yuan
Nominating & Election Committee
Gordon Lan
Publication Committee
Jun Shao
Constitution Committee
Shein-Chung Chow

Current Committee
Membership Committee
Jun Zhao
Awards Committee
Jianping Dong
Communication Committee
Don Sun
Applied Statistics Symposium Committee
Yi Tsong
Book & Journal Donation Committee
Tar Timothy Chen
Annual Meeting Committee
Jiahua Chen
Archive Committee
Ivan Chan
Strategic Committee
Zhiliang Ying

Submission Guidelines
ICSA Bulletin

Articles
The ICSA Bulletin welcomes the submission of articles by the members. Articles submitted should be written in traditional Chinese or English. The contents should be aimed at the general reading.

Articles can be submitted electronically or in papers. If submitted in papers, please leave ¾ inches of blank space on each margin of the letter size paper. The font size of 11 or 12 with Microsoft Word is preferred when typing.

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

Advertisements
The ICSA Bulletin welcomes the submission of advertisements related to the profession of Statistics. The fee for the whole page advertisement is US$300.00. Please contact the Editor for the fee schedule of advertisement with other sizes.

Questions
Please submit your questions to the Editor by email at tsai0123@yahoo.com.

For the complete list of Committee Members and terms, please go to the ICSA website: www.icsa.org.

Letter-to-the-Editor

[Email from a reader]

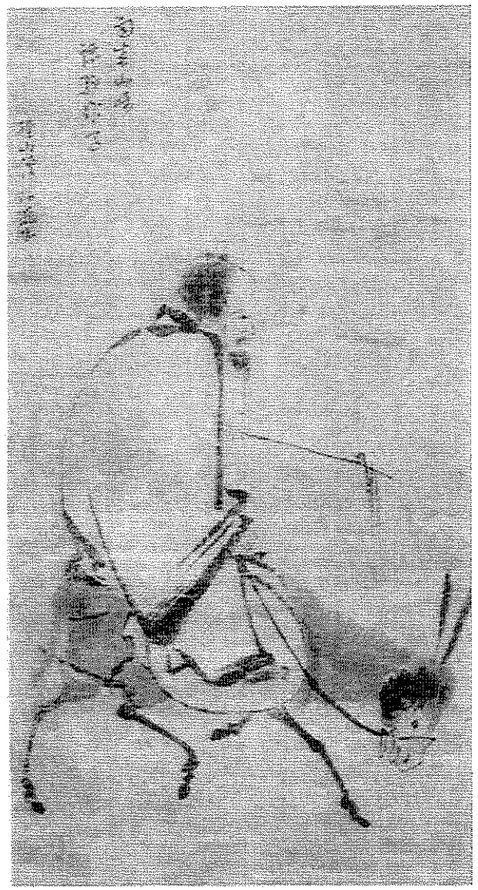
It is interesting to read the "Controversial Statistical Issues." However, the Editor could have better edited the articles to make them more readable. Frankly speaking, many sentences were very difficult to understand due to its length or grammatical errors. I could barely finish reading the first article.

[Reply from the Editor]

Thank you for the precious input. The contributed articles are usually reviewed by my colleagues or friends with English as the first language. Possible revisions or corrections were **suggested** and provided to the authors. It is up to the authors to accept or reject the suggested revisions.

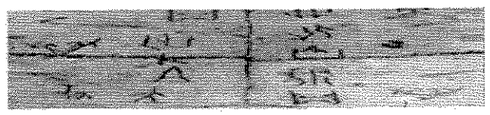
[Email from a reader]

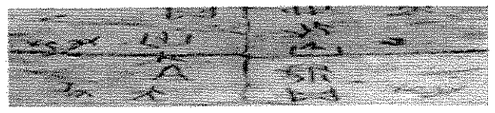
Only two committee Chairs reported the activities of their committees in the last issue of the Bulletin. ICSA seems to have ten committees according to the information in the Bulletin. I would appreciate it if rest of the committees can inform the ICSA members of their accomplishments and the services they provided to the ICSA.



From the Editor:

I would like to thank the readers who spent time to provide us with valuable suggestions. The quality of the Bulletin cannot be improved unless we have more constructive critics. Please write to us again. We hope to hear more opinions, suggestions, or critics from the readers in the future issues.





From the President

Frank Shen, Ph.D.

Mid-year Report: Streamlined Business

Our Symposium

After two years of preparation, the 2004 ICSA Applied Symposium at San Diego proved to be another successful meeting with many milestones. In addition to a record attendance of close to 250 attendees, it was the first time the conference was held on the west coast, during the week instead of weekend, and with an extended three and half day program. The Program and Local Arrangement Committees were cheered for their phenomenal efforts under Drs. Nancy Lo and Gang Li's leadership. We thank their dedicated services to the ICSA.

Our Members

At the membership meeting held at the symposium, we recognized two ICSA members with Distinguished Service Awards. The first recipient was Dr. Sue-Jane Wang. Dr. Wang is currently a senior mathematical statistician in FDA and was the Editor-in-Chief of ICSA Bulletin, 2000-2002. She transformed our Bulletin into a new standard and infused it with a new look through her high vision and extra efforts during her tenure. She is currently one of our Board of Directors and continues to serve in many ICSA committees. The second recipient was Professor Don Sylvekesar. Dr. Sylvekesar is a Professor Emeritus at Statistics Department UCLA with expertise in law and statistics. He is in his third term (i.e., the eighth year) as the Consulting Editor of Statistica Sinica and quietly devoted his time and knowledge to the benefit of our journal and its members. His service to the journal has been invaluable in improving the presentation and clarity of papers that appears in Statistica Sinica. The editing goes far beyond checking for grammatical errors, involving major editorial style changes. We thank their honorable service to the ICSA and congratulate them on the achievements.

We also held a brief memorial service to honor a distinguished ICSA member, Dr. Jiann-Ping Hsu, who passed away early this year with cancer. Her husband, Dr. Karl Peace from Georgia Southern University, donated $10,000 to ICSA and established a Jiann-Ping Hsu (JPH) Memorial Fund with ICSA. ICSA Biometrics section will use this fund to set up a JPH session at each annual applied symposium with a special topic in pharmaceutical and regulatory sciences as well as present a JPH student award. ICSA members who would like to make a contribution to this fund, please send your check to our treasurer in the following address and specify "JPH Fund":

Weiying Yuan, Ph.D.
Director, Biostatistics, J&J PRD
Mail Drop 94, Office: E22800
1125 Trenton-Harbourton Road
Titusville, NJ 08560

Our Board

The Board of Directors had a long but successful meeting as well. Many issues in current business and future direction were discussed. Two topics in particular, special membership fee structure for statisticians in developing countries and decentralizing ICSA to establish chapters or regions, were visited. The former policy has recently been adopted by the American Statistical Association (ASA). We will follow up with both topics in our next Board meeting at JSM in August. If you would like to express your comments in either topic, please feel free to send them to me by e-mail. Other key topics on the August Board meeting agenda will include a contract renewal of Statistica Sinica and the 2007 International Conference site.

Our Committees

Our standing and current Committees are well in place and fully functional in executing annual objectives. We retain most of current committee chairs, with the exceptions that we name Dr. Gordon Lan as the new Nomination Committee Chair and we discontinued Fundraising and Public Relations committees since their roles have been

largely absorbed by the Symposium and Membership Committees. The Nomination Committee has also worked very hard and selected a fine list of candidates for our future leaders. Please take a moment to read the descriptions of candidates and their statements. It is important that you exercise your rights and responsibilities by casting your ballots.

This year's JSM in Toronto is only a few weeks away. Our Executive Director, Dr. Ivan Chan, and colleagues in the Toronto area led by Professor Jiahua Chen of University of Waterloo are working hard on local arrangements, including the ICSA booth, Board and membership meetings, and a Chinese banquet on Wednesday evening. If you are to attend the JSM, be sure to stop by the booth and purchase your banquet tickets as well as meet with old and new friends. Another event, the sixth ICSA International Conference, will be held at the National University of Singapore from July 21 to 23. I wish conference the best and hope that you are one of the few lucky ones that will have a lot of fun in Singapore.

In Closing

I am proud of where ICSA is today. Without a doubt, we are demonstrating tremendous strength in growth. Despite this progress, we need to work hard on diversifying in order to benefit from many exciting opportunities of our time. Diversification includes expanding beyond a single, Chinese ethnic group, reaching out beyond academia, and growing outside North America. My sincere thanks to our members for your continuous support, which has helped us come this far this fast. Our passion is to continuously improve ICSA, and we have only just begun. I look forward to seeing you at JSM/ASA in August.

Frank Shen, President, ICSA

Frank Shen, Ph.D., is Executive Director of Exploratory Development, Global Biostatistics and Programming, Bristol-Myers Squibb Co.
Email: frank.shen@bms.com

From the Executive Director

Ivan S.F. Chan, Ph.D.

I started to write this short note when I was on the flight from San Diego to Philadelphia after attending the 2004 ICSA Applied Statistics Symposium. After two years of preparation of the Symposium Program Committee under the leadership of Nancy Lo and Gang Li, the first ever symposium in the west coast turned out to be a great success with a very strong program and a record number of attendees (about 250). We sincerely thank the Program Committee for their dedication and tireless efforts.

The first time I knew of ICSA was in 1990 when I was a graduate student in Hong Kong and helped with the local arrangements of first ICSA International Conference. Through participation in ICSA activities over the years, I have witnessed a tremendous growth of the organization. This year I am honored to be appointed as Executive Director and to have the opportunity to work with a group of very talented people to serve ICSA.

To date we have records of over 1500 members in our database, but information on some members may be out of date. 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 (Chair of Membership Committee, e-mail: J.Zhao@organonusa.com) or me (e-mail: Ivan_Chan@Merck.Com).

I believe the membership service is essential to keeping our members informed and connected. I would like to thank Jun Zhao for his hard work of maintaining the membership database. One of the important tasks in front of us is to improve the functionality of the membership directory and the web site in order to make them more easily assessable and useful for our members. The ICSA Board of Directors discussed this issue at its meeting on June 6 at the ICSA Symposium and commissioned a task force to work on a solution. The task force will include Frank Shen (President), Jun Zhao, Don Sun (Communication Committee Chair), Yi Tsong (Past Executive Director), Naitee Ting (Program Committee Chair), and me. As always, we welcome your suggestions.

This year we will elect several officers, including 2005 President-Elect, 2005 Biometrics Section Chair, and 6 Board of Directors (2005-2007 term). I would like to thank the Nomination Committee, under the leadership of Gordon Lan, 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 reviewed at the board meeting at the JSM in Toronto and announced at the Annual Members Meeting on August 11, 2004.

If you are planning to attend the Joint Statistical Meetings (JSM) in Toronto (August 8-12), 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 11 (Wednesday, 6:00-7:00pm, Metro Toronto Convention Center, Room 707). Continuing our tradition, we have arranged a dinner at an authentic Chinese restaurant following the Annual Members Meeting. Dinner tickets are available for sale at the ICSA booth. We sincerely thank Jiahua Chen (President-elect) and members of the Annual Meeting Committee for coordinating these important activities.

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

Ivan S. F. Chan, Ph.D. is Director of Clinical Biostatistics at Merck Research Laboratories.
Email: Ivan_Chan@Merck.Com

Reports From Committee Chairs

By: Naitee Ting, Ph.D. & Jun Zhao, Ph.D.

Program Committee

By: Naitee Ting, Ph.D.

As I am drafting this report, the 2004 symposium committee members are busy preparing for the June Applied Statistics Symposium. This committee (chaired by Nancy Lo and Gang Li) has been working very hard for the symposium since many years ago. As you may recall, they put all of the efforts together to prepare the 2003 ICSA Applied Statistics Symposium, but that symposium was postponed because of SARS. For this decision, the committee members had to notify all meeting participants, refund the fees, withdraw the facility reservations, ... and perform many other very difficult tasks. This year, the same committee is working hard again to prepare for the 2004 symposium. On behalf of the Program Committee, I would like to take this opportunity to express our sincere appreciation for their effort, and I hope you can also let the symposium committee members know that we are grateful of their help and their hard work.

I look forward to meeting you at the June Symposium in San Diego.

Meanwhile, another committee is also working very hard to prepare for the upcoming JSM activities for ICSA. The annual meeting program committee (chaired by Jiahua Chen) identified a Chinese restaurant in Toronto, and committee members will help set up the ICSA booth at JSM. As you know, each year, we need members to help staff the ICSA booth. If you will be available to help with the ICSA booth, please notify the Executive Director, Ivan Chan (email: ivan_chan@merck.com). He is in the process of setting the ICSA booth schedule. The announcement of the JSM activities (membership meeting on Wednesday afternoon and dinner banquet on Wednesday evening) can be found from this issue of the ICSA Bulletin.

The Sixth ICSA International Conference will be held in the National University of Singapore, July 21-23, 2004. The organizing committees are currently working on the scientific program as well as local logistics. Updated information can be found at the ICSA website (http://www.icsa.org/) and the NUS website (http://www.statistics.nus.edu.sg/).

The 2005 Applied Statistics Symposium will take place in Washington, D.C. Details regarding this symposium will be announced in the near future.

The International Conference on Multiple Comparisons (MCP) plans their next meeting at shanghai, China, and they are interested in co-sponsorship with ICSA. Some ICSA members have been involved in the planning stage and that the MCP committee has made important progress already. The MCP will be in 2005, and the committee is chaired by Professor Jason Hsu.

The program committee proposes to hold the 2006 Applied Statistics Symposium at Connecticut. Details will be worked out in the near future. This proposal will be discussed in the June Board meeting at San Diego.

The program committee proposes to hold the next ICSA International Conference at Taiwan. This proposal will be discussed in the August Board meeting at Toronto.

Naitee Ting, Ph.D. is an Associate Director in Biostatistics, Pfizer Global Research & Development, New London, CT 06320, USA
Email: naitee_ting@groton.pfizer.com.

Membership Committee

By: Jun Zhao, Ph.D.

ICSA is getting more and more popular in Statistics community. According to current ICSA membership database, we have enrolled overall more than 1500 past or current members. And there are more than 400 active members who are either permanent members or who have paid current year membership fees.

One benefit from ICSA is that each active member can receive two issues per year of unique and wonderful ICSA Bulletins edited by dedicated Bulletin committee. The Bulletin is one channel of communication between members and the organization. Other communication tools include telephone, e-mail, website (www.icsa.org) and regular mail. Since the members' contact information depends largely on the data in the current membership database. Therefore, it is important to maintain an accurate database. Thus one important objective for the current membership committee is to maintain and update the membership database. The ICSA executive team and the membership committee have given great effect on the database and are still working very hard on them, with helps from some other committees.

Meanwhile, an email distribution list has been used according to the most recently updated information from the ICSA membership database. In order to serve our members more efficiently and to reduce time and cost spend from the ICSA, we encourage members (especially permanent members) to update contact information periodically.

Currently ICSA members have the following options to update their contact information: 1) Mail the membership update forms to Dr. Ivan Chan; 2) Email update information to Ivan Chan (ivan_chan@merck.com) or Jun Zhao (j.zhao@organonusa.com); 3) Get a Login Name and a Password from Ivan either Chan or Jun Zhao, then update contact information online through membership section of the ICSA website (www.icsa.org); 4) Contact local ICSA membership committee members: Ling Chen, Rongdean Chen, Ming Tan, Heping Zhang [US]; Wai-sum Chan [HK]; Qiwei Yao [UK]; Chen-Hsin Chen [TW]; Guo-Ying Li [CN]; Jiahua Chen [CA].

For the recruitment of new members to the ICSA, we need supports from all ICSA members. One way is to expose the ICSA organization more to universities, other societies and some areas such as China. And this needs support from current members, especially who are from universities. We encourage you to introduce our organization to your students, your colleagues and your friends.

Jun Zhao, Ph.D. is a biostatistician at Organon Pharmaceuticals USA Inc, Roseland, NJ, USA.
Email: j.zhao@organonusa.com

Report From Symposium Committee Chairs

Nancy Lo, Ph.D. & Gang Li, Ph.D.

The ICSA 2004 Applied Statistics Symposium was held from June 6 – 9 at San Diego Marriott Hotel La Jolla. The theme of the symposium was 'Statistics in Bio-tech Research and Computing Intensive Methodologies'. We invited Professor Bradley Efron, Max H. Stein Professor of Humanities and Sciences, Stanford University and Professor George Tiao, W. Allen Wallis Professor of Econometrics and Statistics, University of Chicago as two keynote speakers. Their talks on 'Bayesians, Frequentists and Microbiologists' and ' The Split and Recombine Procedures: 'A Diagnostic Analysis of Heterogeneous Data', although on different aspects of application of statistics, are both innovative and informative. Attendees listened to two plenary session speakers: Professor Jiangqing Fan, Princeton University and Professor Tze Leung Lai, Stanford University on Tuesday, June 8th on 'Normalization and Significant Analysis of Cdna Micro-arrays using Within-array Replications' and 'Current Statistical Issue in Clinical Trials for Drug Development'.

Five short courses were offered: "Practical Guidance of Generalized Linear Mixed Models" by Charles E. McCulloch, University of California, San Francisco; "Tutorial on Statistical Bioinformatics" by Jun Liu, Harvard University; "Resampling Methods; A guide for Practitioners" (a change of instructor for this short course, due to illness of Mike Chernick) by Dr. Phil Good, Huntington Beach CA; "Cancer Trials for Practitioners – Experimental Design and Efficacy Analysis" by Kao-Tai Tsai, Aventis Pharmaceuticals; and "Active Controlled Clinical Trials" by Yi Tsong & Sue-Jane Wang, FDA. We had 84 people registered for the short courses on June 6, Sunday, prior to the technical sessions.

In addition to two keynote speeches and two plenary sessions, we had nine sets of four concurrent technical sessions including five sessions of contributed papers during June 7-9 2004. Among the topics were: Statistical methods in assessing agreement, Non-stationary continuous-time models, Estimating predictors for long-or-short-term survivors, Computing marginal likelihoods from a signal MCMC output, On the superiority of adaptive designs, On identifying differentially expressed gene from DNA microarray data, Single variable control charts: An overview, Spatial bootstrap and On the current status of capture-recapture models in wildlife research, just to name a few. One special feature at this symposium was the first invited session entitled "Jiann-Ping Hsu Pharmaceutical and Regulatory Sciences".

The late Dr. Jiann-Ping (JP) Hsu, a long time dedicated ICSA member, sadly lost her three-year battle with breast cancer on February 9, 2004. Dr. Karl Peace, her husband, in memory of Dr. JP Hsu, generously donated $10,000.00 to ICSA to establish an endowment fund: "Jiann-Ping Hsu Memorial Fund". Moreover, since the late Dr. Hsu had a distinguished career in Pharmaceutical industry and high impact on regulatory sciences, ICSA will establish a special student paper award, entitled "Jiann-Ping Hsu Pharmaceutical and Regulatory Sciences Student Paper Award," to encourage innovative research in this area. Starting 2005, the award will be selected by the annual symposium committee along with other student travel awards and receive the same monetary sum as the other student travel awards.

The traditional Monday night banquet, June 7th was held at Kings Garden Chinese Restaurant, 3.5 miles from the hotel. A total of 200 attendees and their family members were present. As people entered the restaurant, they were greeted by the tranquil Chinese music of Yangqin, Chinese hammered dulcimer, performed by Shilin Wang, Statprobe dressed in her traditional Chinese cheepau. Naitee Ting was M.C. for that evening. Distinguished guests included Professor Chung-Kuan Cheng, Department of Computer Science, UCSD and Dr. Iwen Yao, Qualcomm, the President and VP of San Diego Chinese American Science and Engineering Association. Near the end of the 10-course cuisine, Professor Arlene S. Ash, Boston University gave a dynamic talk on

'Statistics as if it matters: why not change the world?'

Then the belated surprise celebration of Professor George Tiao's 70th birthday (supposed to be 2003), moderated by Xiaoli Meng with photos of George from his early years as a young professor to this day. The slide show brought much laughter especially when the thin-faced George showed up with his bride on their wedding day. George gave a brief account of how the initial group that eventually evolved into ICSA gathered periodically in the late 1960's in their basement where he and his wife entertained this group of young graduate students in statistics at University of Wisconsin. So 30+ years later, the fact that the ICSA celebrated his 70th birthday in San Diego CA was unthinkable to Professor and Mrs. George Tiao. To show our appreciation toward Professor Tiao's leadership, Dr. Frank Shen, the president of ICSA, presented George an Oscar type of trophy. As some were leaving the restaurant, most stayed for the Karaoke, led by Yi Tsong and Ms. Sunny Lei, a well known singing star in San Diego Chinese communities, while David Shen handled the karaoke machines. People enjoyed singing their favorite songs, among whom was Lawrence Lin, singing just like Elvis Presley, till 10:30pm, the predetermined time for the buses to bring people back to the hotel.

2004 was the first time an ICSA symposium was held in the west coast. The symposium lasted for three and half days from June 6-9 (Sunday to Wednesday noon) 2004. There were close to 270 attendees, including attendees from China, Taiwan, and England. Approximate 115 attendees were from California, 3 from Oregon and 4 from Washington State. This symposium was originally scheduled for June 2003, but was postponed to 2004 due to the SARS crisis in April-May, 2003. The postponement of the 2003 symposium put an extra burden on both logistic and program committee members. We were able to keep the program intact, and Kathy Chi-Burris, our treasurer, dealt with reimbursement of all fees for 2003 and answering all the questions. We commended our committee members for their dedication, forbearing and persistence: program committee: Gang Li, Naihua Duan, Keh Shin Lii, Joey C. D. Lin, Ying Lu, Kung Jong Lui, Edward Pun, Weng-Kee Wong, and Eric Yan; Logistic Committee: Nancy Lo, Alice Chu, Kathy Chi-Burris, William Yuan, David Shen, George Yu, John Lee, Thomas Lin, Kung Jong Lui, Xun Lin, Christina Show, Eric Yan, Jenny Han, Joey C. D. Lin, Edward Pun, Ming Ji, Yan Wang, Feng He, Juanjuan Fan, Oralia Loza, Lin Liu and Shilin Wang. We thank all session organizors who helped putting together an excellent symposium program. We appreciate the understanding of all the speakers, who stuck to their commitments to presenting their talks at the 2004 symposium. We thank ICSA executive committee, in particular, Zhanliang Ying, the past president, Yi Tsong, the past executive director, and Don Sun, our ICSA webmaster, who shared the burden during that period. Over 30 students from SDSU and UCLA assisted with registration and the visual equipments.

The Southwest Fisheries Science Center, National Marine Fisheries Service, NOAA, La Jolla provided great support in lending visual equipments for presentations, design of registration procedure and symposium flyers and posters. This symposium would not have been possible without generous donations from the following companies, to whom we express our sincere appreciation: Allergan Inc., Amgen Inc., Aventis Pharmaceuticals Inc., Biogen Idec Inc., Bristol-Myers Squibb Co., Hoffmann-La Roche Ltd., Johnson & Johnson, Merck & Co. Inc., Pfizer, Inc., ProSanos Corporation, Statplus Inc., and Valeant Pharmaceuticals International.

Last but not the least, we want to thank all speakers and participants, without whom, ICSA 2004 symposium would not have been a realty. This ICSA 2004 symposium in San Diego is indeed a sign of growth and geographic expansion of ICSA and we hope that future ICSA symposium meets in every part of the U.S.A.

Nancy C. H. Lo, Ph.D. is a mathematical statistician of Southwest Fisheries Science Center, La Jolla, CA
Email: Nancy.Lo@NOAA.GOV

Gan Li, Ph.D. is a professor of Statistics, University of California, Los Angeles, CA.
Email: gli_la@yahoo.com

Jiann-Ping Hsu Memorial Session

By: Karl E. Peace, Ph.D.

Thank you very much for the opportunity to say a few words about JP and our relationship, to this audience, which includes so many of her friends, colleagues and fellow professionals. My comments will highlight some of her accomplishments and her many virtues, from the time I first knew her.

I first knew the name Jiann-Ping (**JP**) Hsu in the Fall of 1982 when I went to Smith Kline & French Labs (SK&F). SK&F Labs had an oral gold compound under review by the FDA at that time, specifically by John Harter and JP. The lead statistician for the project reported to me and often had meetings with John and JP, and would bring back requests from JP and John regarding additional analyses, for example "Q statistics", which was Dr. Harter's acronym for confidence or rather, fiducial intervals arising from Fieller's theorem. I then met JP for the first time, in 1983, in Toronto, following a session on positive controlled trials (yes that was the rubric at that time, later active controlled equivalence trials (**ACES**) became the phrase, and currently non-inferiority trials is in vogue) at the Joint Statistical Meetings, in which she, Satya Dubey, Kathleen Lamborn and Lloyd Fisher, participated. I can still picture her during her presentation often pushing back her beautiful hair from her face.

In 1986, JP joined SK&F Labs, managing a group of statisticians providing statistical support to the non-clinical areas of R&D. She spent two years at SK&F labs, performing her responsibilities superbly. Thereafter she accepted a management position in biostatistics, at Parke-Davis/ Warner Lambert, in Ann Arbor, where again she made many important contributions. A few years later, she accepted the position of Associate Director at Schering Plough Research Institute, where she was very influential on the professional growth of many of her reports. In fact, Kao-Tai Tsai has often remarked to me that "JP is the best boss I ever had." Parenthetically, when these meetings were held at Georgetown University, Dr. Chin Long Chiang, JPs major professor, told me that JP was "a superb human being, and a brilliant student, having solved a difficult problem which was helpful to their research." A portion of the solution to this problem, I believe, was what Dr. Chiang referred to as "a useful lemma" in his Keynote address at the Georgetown meetings. JP was in fact the recipient of the Evelyn Fix award at UC Berkeley, which was awarded to the student demonstrating the greatest promise for quality research.

From Schering, JP moved to Norwich-Eaton, a subsidiary of Proctor & Gamble, as Head of Biostatistics and Data Management. In 1992, she joined Biopharmaceutical Research Consultants, Inc. (**BRCI**), in Ann Arbor, Michigan, as Director of Operations. BRCI is a full service contract research organization **(CRO).** A year thereafter, she was promoted to Vice President of Operations.

JP and I were married June 03, 1993, a short 11 years ago. In the late 1990s, JP became President and CEO of BRCI, in charge of all business and operations. JP continued to grow BRCI, securing contracts with several pharmaceutical companies and with the National Institute of Drug Abuse (**NIDA**), at the NIH. Some of her contributions to NIDA were instrumental in the development of their Clinical Trials Network (CTN). Many top level individuals at NIDA have often remarked to me, what a superb person, and professional, JP was. Her contributions were greatly appreciated. In fact, everywhere JP worked, she made lasting friends and important contributions. She did this

in her own genteel, graceful, non-aggressive and non-confrontational style. She was a lady, whose entire personna, commanded respect.

Over the years, JP was active within the Drug Information Association (**DIA**), the Pharmaceutical Manufacturers Association (**PMA**), the American Statistical Association (**ASA**), particularly The Committee on Women in Statistics (**COWIS**), the Biopharmaceutical Applied Statistics Symposium (**BASS**), and the International Chinese Statistical Association (**ICSA**). All professional activities were important to her, but most of all, she treasured those within the ICSA. Naitee will share specifics of JPs ICSA activities in just a moment. I know were JP to send a message to this meeting, she would wish her many ICSA friends well, and want that they stay true to the many virtues she demonstrated.

Although she would never draw attention to herself, I know that she would be honored by my activities to disseminate to a wide audience, knowledge of her stellar qualities. These activities include The Jiann-Ping Hsu Award for Excellence and Scholarship at UC Berkeley, The Jiann-Ping Hsu School of Public Health (**JPHSPH**) at Georgia Southern University (**GSU**), which is the first school of Public Health in the entire University System of Georgia, and the permanent establishment of the Jiann-Ping Hsu Regulatory and Biopharmaceutical Sciences Session (**JPHRBSS**) at the annual meeting of the ICSA. The endowment creating the JPHRBSS will also provide a modest honorarium to the author of the outstanding research paper presented at the JPHRBSS. Regarding the JPHSPH at GSU, the plaque permanently mounted in the JPHSPH at its dedication in late January of this year, includes a picture of JP and a description of her accomplishments and virtues. One quote from the plaque reads: "JPs zeal for excellence, consideration of others, intelligence and scholarship, honesty and kindness", and the "distinguished service in her life's work, and devotion to quality and caring, in all her endeavors personify the touchstones of learning and leadership in the health professions."

At a personal level, it has been my greatest joy to have been a part of her life, and to contribute to her professional growth. I know that I will always miss her, admire her, love her, and remain thankful for the almost 11 years we had as husband and wife. Not only was she my wife, she was my best friend, my confidant, and the person for whom I had the most admiration. She was indeed my soul mate. My fondest wish is that JP and I will be reunited in the afterlife.

Penultimately, JPs delicateness, gentleness, beauty, grace, and her appreciation of my fondness for poetry and verse, reminds me to close with the following:

A Butterfly lights beside us
Like a burst of light
For A brief Moment
Her Glory, Grace and Beauty Belong to Our World
But Then She Flies on Again
Although We wish She could have Stayed Longer
We are so glad, and feel so Fortunate
That She chose to spend some Time with us.

Finally, I am happy to present this check in the amount of $10,000.00 to Frank Shen, current President of ICSA, as the initial seed establishing the JPHRBSS at ICSA.
Thank you.

Jiann-Ping Hsu Memorial Fund

By: Naitee Ting, Ph.D.

It is a great honor for me to take this opportunity to say a few words about Dr. Jiann-Ping (JP) Hsu, a good friend to ICSA, and a role model for many applied statisticians. Dr. Karl Peace, JP's husband, talked about her successful life story, and a great deal of her contributions to the field of statistical applications both in pharmaceutical research and development and in regulatory sciences. I will cover some of her contributions to ICSA, and how she positively influenced many people, including myself.

In addition to her efforts devoted to many professional societies including ASA, DIA, PMA, COWIS, and BASS, JP also made lots of major contributions to ICSA. Below is a list of important roles she played in our association:

- **Member of Board of Directors**; 1991-1993, 1997-1999.
- **Chair of the Biometrics Section**; 1994.
- **Co-chair of the Program Committee** for ICSA 2001 Applied Statistics Symposium.
- JP received the **Meritorious Service Award** from ICSA in 1995. (Statement: "Awarded for distinguished service, dedicated effort and unselfish support").

Every time when there was an ICSA Applied Statistics Symposium, JP made donations. She also donated to other ICSA activities. In the late 1980's and early 1990's, JP helped promoting applied statistics within ICSA. She also motivated many members to be involved in the ICSA activities.

Although JP made many important contributions to ICSA, some of our members may have never heard of her. This is because JP had a soft personality, and was very low key in working with people. JP was a dedicated, hard and productive worker. She set high standards on herself, but was very nice and kind to other people. Her personality was consistent with a Chinese proverb: "Ask a lot of herself, but ask very little of other people".

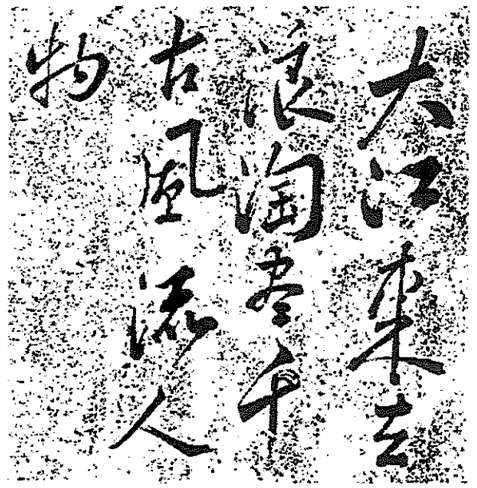
In JP's family, she was the eldest among the 5 daughters. JP was also the first in her family to arrive at U.S. and to establish herself in this country. She then brought her parents and sisters to America, and provided for their care and college education for her sisters.

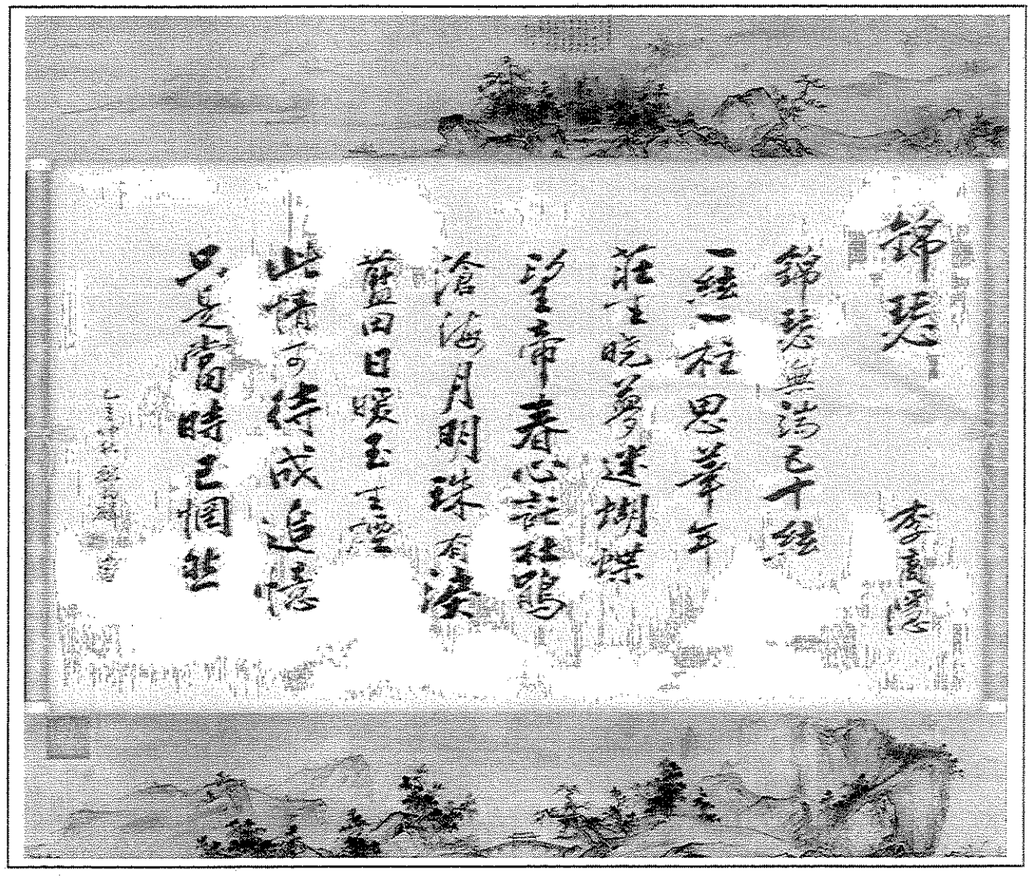
In her professional life, JP was able to successfully mentor more than 60 biostatisticians employed in pharmaceutical industry or government positions. This is a tremendous achievement because mentoring takes a lot of patience and caring. With her patience and caring nature, JP was easily among one of the most successful mentors.

In the months before passing away, JP fought courageously against the cancer. Throughout the whole process, she was very optimistic, and she maintained her low key approach and not letting many people know her disease. When some of the very close friends asked her to share some of her pain with other friends, she softly refused by saying "If I tell them, they can't help, but they begin to worry about me". - She was still very thoughtful in the later stage of her life.

In memory of Dr. JP Hsu, ICSA establishes a "Jiann-Ping Hsu Memorial Fund" with the endowment of $10,000, as the fund seed from Karl Peace. ICSA will also establish a special student paper award, entitled "Jiann-Ping Hsu Pharmaceutical and Regulatory Sciences Student Paper Award," to encourage innovative research in this area. Starting 2005, the award will be selected by the annual symposium committee along with other student travel awards and receive the same monetary sum as the other student travel awards. Furthermore, ICSA will also establish a regular invited session entitled "Jiann-Ping Hsu Pharmaceutical and Regulatory Sciences" in the annual ICSA Applied Statistics Symposium. The recipient of the JPH Award will present the award paper along with other invited speakers. The recipient will also receive a plaque commemorating the event. ICSA will publicize the award and the fund to encourage public contributions that will enrich the fund.

In my personal opinion, it is very appropriate to use this fund to award students. I believe every student should consider JP as a role model from both professional and personal point of view. Eventually, students who commemorate JP and follow her way of life will be successful and prosperous in every aspect.





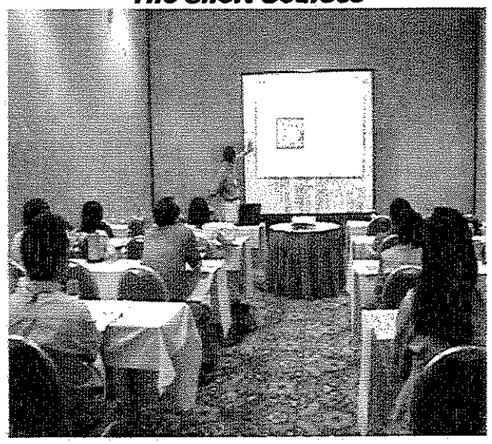
ICSA 2004 Applied Statistics Symposium

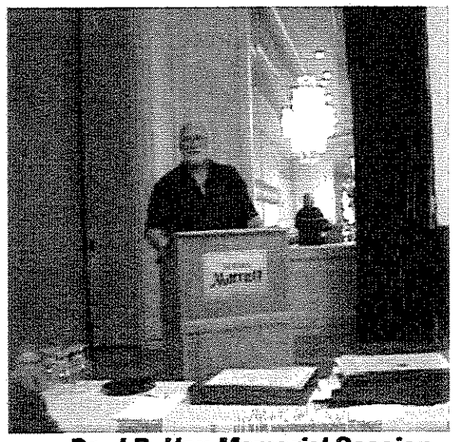
Photographic Highlights

Supplied By Symposium Committee

Symposium Organizers and ICSA Members

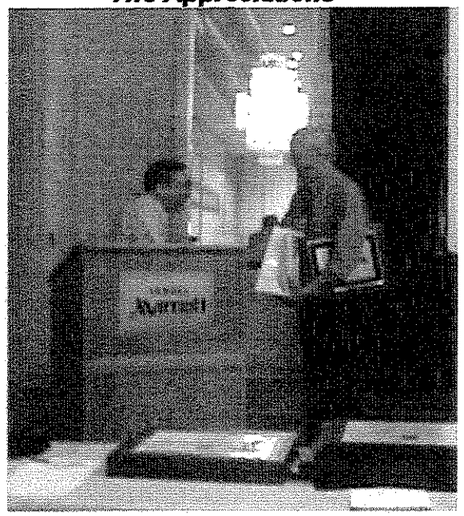
The Short Courses





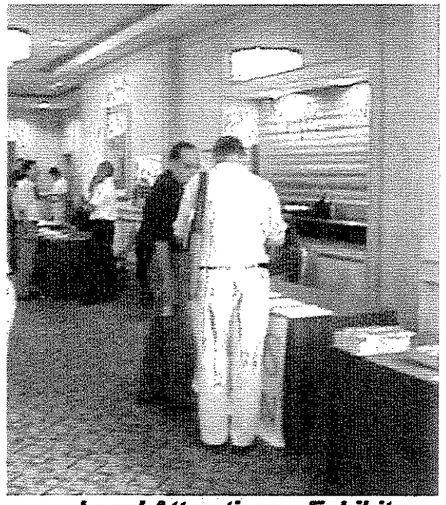
Dr. J.P. Hsu Memorial Session

The Appreciations



Statistical Interactions

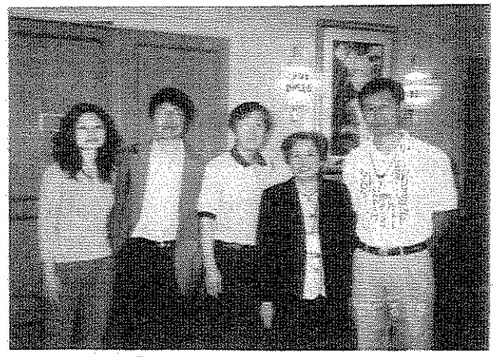
Student Awards



Local Attractions - Exhibits

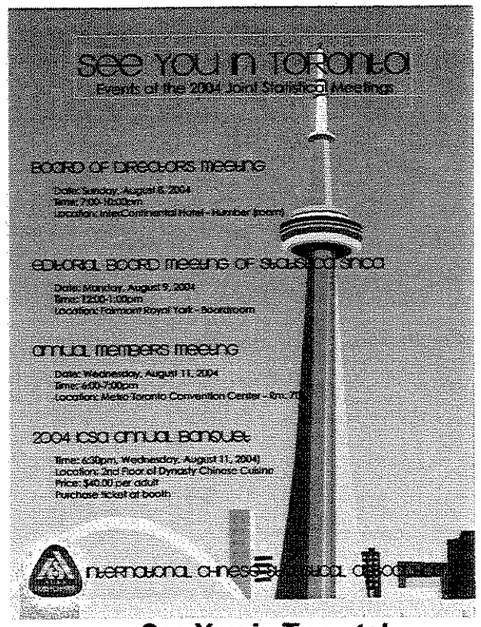
The Appreciations

Representatives of the Generous Corporate Sponsors



Committee Members

The Unknown Heroes



See You in Toronto!

Congratulations & Thanks!

From the Editors, Statistica Sinica

Hwai-Chung Ho, Ph.D. & Jane-Ling Wang, Ph.D.

Another year has passed since our last column on the prospect of the journal ***Statistica Sinica***. With the support of the Associate Editors, the referees and every author, we are pleased that ***Statistica Sinica*** has continued making progress, steadily moving on to raise its standard as a world-leading journal.

The total number of papers submitted in 2003 was 219, up about 20% from previous years. Out of the 193 submissions received from June 1, 2003 to April 30, 2004, the first-round reviews had been completed for 121 of them, and 23 (19 percent) of them were given the status of tentatively accepted or accepted. This acceptance rate is lower than the average of the previous years, and signals an increase in the journal standard. We have also received more papers in the general area of biostatistics with more diverse topics. This is reflected partly in the special issue on "Emerging Issues in Longitudinal and Functional Data Analysis," to appear in the July issue of 2004. Four emerging areas are included: nonparametric and semiparametric, joint modeling of longitudinal and survival outcomes, methods for handling missing data, and approaches to making casual inference. The incorporated 18 papers organize a rich and interesting theme, upholding a milestone of the frontier research on Longitudinal Data Analysis.

Another special theme issue, which was planned at the beginning of 2003, is going smoothly. This issue entitled "Bayesian Inference, Environmental Statistics, Time Series Analysis, and their Applications," similarly takes in many promising and critical papers, is scheduled to be issued in January 2005. Also on the theme topics front, a cluster of eight papers on "Statistical Applications in Financial Econometrics", appeared in the October issue of 2003.

Following our goal to accelerate the review process, the median time to first review has been about four months, and to a full review about six months (this excludes the time spent by authors on revisions). A few papers took much longer than usual, and we have been monitoring closer to avoid such long delay in the future. For the 656 subscribers of the latest issue (April 2004), 417 are the members of **International Chinese Statistics Association (ICSA).** Furthermore, the online web (http://www.stat.sinica.edu.tw/statistica/) is now extensively used. We continue trying to make the interface user-friendly and the search functions complete, hoping that the reviewers, as well as the readers, would enjoy the convenience of utilizing the web.

The writing of articles, on which we put lots of emphasis, has gradually gained ground according to Dr. Don Ylvisaker, who helps ***Statistica Sinica*** to review the editorial content of each accepted manuscript. Since 1998 when Don received the appointment as the Consulting Editor, he has played the most important and indispensable role in maintaining high-quality presentation of each paper appeared in ***Statistica Sinica***. To thank Don for the enormous contribution he has made to the journal, the ICSA presented a Distinguished Service Award to Don during the ICSA Applied Statistics Symposium (June 6-9, 2004) held in San Diego, California. Dr. Ching-Shui Cheng, on behalf of the Institute of Statistical Science of Academia Sinica, also presented a gift at the Symposium to Don to jointly show our deepest appreciation for his outstanding work.

According to the latest SCI Journal Citation Reports published by the Institute for Scientific Information, in 2002 the total cites of ***Statistica Sinica*** grew into 525 from 420, with the Impact Factor increasing to 0.605 from 0.467, compared with 2001. We thank all the reviewers and the authors, without whom we could have not accomplished this. We also appreciate the generous support from members of **ICSA**. We expect that, with your constant participation and sponsorship and with the past experience of refining the journal, ***Statistica Sinica*** would reach another acme of success in the immediate future.

Hwai-Chung Ho, Ph.D., is Research Fellow of the Institute of Statistical Science, Academia Sinica, Taipei, Taiwan, R.O.C.
Email: hcho@stat.sinica.edu.tw

Jane-Ling Wang, Ph.D., is Professor of Statistics, University of California, Davis, CA, U.S.A.
Email: wang@wald.ucdavis.edu

STATISTICA SINICA

Volume 14 Number 3 July 2004

Emerging Issues in Longitudinal and Functional Data Analysis

Non and Semiparametric Regression

Joint Modeling of Longitudinal and Time-to-event Data

Causal Inference

Missing Data

Gerenal

Candidates for 2004 Election of ICSA Officers

President Elect - 2005

Biometrics Section Chair - 2005

Board of Directors (6) - 2005 to 2007

Candidates for 2005 President Elect

LEE, Mei-Ling Ting

[PRESENT POSITION] Associate Professor in Medicine (Biostatistics), Harvard Medical School; Associate Professor in the Department of Biostatistics, Harvard School of Public Health; Biostatistician, Brigham and Women's Hospital, Boston, Massachusetts, USA. **[FORMER POSITION]** Assistant Professor, Boston University, 1984-1992; Statistical Coordinator, Data General Corporation, 1981-1983. **[DEGREES]** B.S. in Mathematics, National Taiwan University; M.S. in Mathematics, National Tsing-Hua University, Hsin-Chu, Taiwan; Ph.D. in Mathematics/Statistics, University of Pittsburgh, 1980. **[FIELD OF MAJOR STATISTICAL ACTIVITIES]** Statistical Methods for Microarray Gene Expression and Proteomic Patterns, Statistical Applications in Microbiology and Pharmacokinetics, Epidemiological Research and Nonparametric Methods for Clustered Data, Lifetime Data Analysis and Models for Latent Health Status, Statistical Methods for Cancer and Genetic Research **[SELECTED PUBLICATIONS]** Recently, Dr. Lee has written a book titled *Analysis of Microarray Gene Expression Data* (2004) published by Kluwer Academic Publishers. Dr. Lee also has two edited volumes: Mesbah M, Cole BF, Lee MLT, editors, *Measurements and Statistical Analysis for Quality of Life Data* (2002), Kluwer Academic Publishers; Jewell NP, Kimber AC, Lee MLT, Whitmore GA, editors, *Lifetime Data: Models in Reliability and Survival Analysis*, (1995), Kluwer Academic Publishers, Dordrecht, The Netherlands. Dr. Lee has published more than 60 original articles in peer-reviewed international journals including: *Annals of Probability, Proceedings of the National Academy of Sciences,Statistics in Medicine, JRSS Series B, Biostatistics, Journal of Applied Probability, Biometrics, Journal of Infectious Diseases, Annals of Epidemiology,* etc. **[ICSA ACTIVITIES AND OFFICES HELD]** Publication Chair (1997); Member, Board of Directors, International Chinese Statistical Association (2001-2003). **[RELATED PROFESSIONAL ACTIVITIES]** Dr. Lee is the Founding Editor and Editor-in-Chief, *Lifetime Data Analysis* (1994-present). Dr. Lee is an Elected Member, International Statistical Institute (1995), Fellow of the Royal Statistical Society, U.K. (1998), and Fellow of The American Statistical Association (1999), USA. **[STATEMENT]** Being nominated as a presidential candidate for ICSA is a great honor. If elected, I will commit my time and energy and, indeed, my heart and soul, to strengthening the academic, professional and financial aspects of ICSA to the best of my ability. I have shown a long commitment to ICSA through continuous membership and service. I believe strongly in the future importance of ICSA to the Chinese statistical communities worldwide from which it draws its strength. My background and experience are diverse and international in perspective. These give me a unique perspective from which to understand the needs of ICSA and its membership. As a President, I will consult with all of you and work diligently to lead ICSA toward a bright and hopeful future. The following are some of my key aims if elected: (1) To promote *Statistica Sinica* and make it the journal of choice for important statistical advances by using my ten years of experience on directing an international journal; (2) To build membership in ICSA, especially among younger statistical professionals, through promotion of the ICSA and development of programs that increase its relevance and value to potential constituencies; (3) To help the ICSA play a more active role in information exchange concerning professional affairs, placement and research opportunities for its members; (4) To make the bi-annual conference a significant and essential international meeting for all members and the wider statistical community; (5) To promote the ICSA as a

truly international statistical body to which the whole statistical community will look for leadership and direction; (6) To examine the potential role of ICSA in disseminating knowledge and information in our field through the use of electronic media and internet services, including our ICSA website, on-line publishing and data warehousing.

This list of goals is a starting point that will be modified and extended when I have an opportunity to consult with ICSA present and future members, and discover and tackle pressing problems. We will nurture opportunities for ICSA to fulfill its mission. To serve you would be a privilege. Working together we can achieve the great promise that ICSA has always offered and meet the ever-present needs of statistics and statisticians in this rapidly changing world.

TSONG, Yi

[PRESENT POSITION] Dr. Tsong is the Acting Deputy Director of the Quantitative Methods and Research Staff, Office of Biostatistics, Center for Drug Evaluation and Research, U.S. Food and Drug Administration. **[FORMER POSITIONS]** Before joining FDA in 1987, he was a Biostatistician at The University of Texas Medical Branch at Galveston in 1983-1986, a Sr. Scientist at Lockheed Electronic Engineering and Management Service Corp in 1980-1983, and a Post-doc fellow at the Northwestern University Medical School in 1978-1979. **[DEGREES]** He received his Ph.D. in Mathematical Statistics from the University of North Carolina – Chapel Hill in 1979. **[FIELD OF MAJOR STATISTICAL ACTIVITIES]** In the recent five years, his major statistical interests are in pharmacoepidemiology, clinical trial design and analysis and statistical quality control. **[SELECTED PUBLICATIONS]** Dr. Tsong has published more than 50 research articles in books and referred statistical, medical and pharmaceutical journals. Dr. Tsong's recent publications include: "Postmarketing adverse event signaling processes", "Dissolution Profile Similarity" and "Non-inferiority analysis in active controlled clinical trials" in *Encyclopedia of Pharmaceutical Statistics*, 2nd Ed. 2003; "Pre-marketing shelf life determination of a pharmaceutical products", in *Encyclopedia in Pharmaceutical Statistics*, 3rd edition, 2004; "Statistical issues on objective, design and analysis of non-inferiority active-controlled clinical trial", J. of Biopharm. Statist. 2003; "Adaptive Group sequential methods for Bioequivalence assessment of non-systemic drug products", J of Biopharm. Statist. 2004; "Sequential drug dissolution sampling acceptance rules", J. of Biopharm. Statist., 2004; "Significance level of stability pooling tests", J. Biopharm. Statist. 2003; "ANCOVA approach for shelf life analysis of stability study of multiple factor designs". J. Biopharm. Statist. 2003; " Shelf life determination based on equivalence assessment", J. Biopharm. Statist. 2003; **[HONORS AND AWARDS]** Dr. Tsong received ICSA Distinguished Service Award (1996) and ICSA Certificate of Appreciation (2000). He received the following awards for his contributions in FDA: FDA Outstanding Service Award (2001, 2002); FDA Award of Merit (1993, 2001); FDA Scientific Achievement Award (2003); FDA Commendable Service Award (1994, 2000); FDA Group Recognition Award (2002, 2004); CDER Special Recognition Award (2000, 2001), CDER Team Excellence Award (2000, 2001, 2002), CDER Center Director's Special Citation (1998); FDASA Special Act and Service Award (1997). **[ICSA ACTIVITIES AND OFFICES HELD]** Dr. Tsong served in ICSA as Executive Director (2001-2003); Member of ICSA Applied Statistics Symposium Planning Committee (1990, 1993-1996, 1999); Chair of 2005 ICSA Applied Stat. Symposium; Member of Board of Directors (1998-2000); Chair of Biometrics Section (1997); and Treasurer of ICSA Applied Statistics Symposium (1997-1999). **[RELATED PROFESSIONAL ACTIVITIES]** Dr. Tsong is a member of ICSA and ASA. He is co-chairing the first PhRMA/FDA CMC Workshop in 2005. Currently he serves as an Associate Editor of J. of Biopharmaceutical Statistics (2003-2005) and as the coordinator of FDA CDER Office of Biostatistics Seminars since 1996. Dr. Tsong served as a guest editor of J. Biophar. Statist. in 2003. He served as a referee for various statistical and pharmaceutical journals. He served in FDA Statistical Association as the chair of the first FDA Statistics Symposium (1996); in the By-law Committee (1995-1997), as the seminar coordinator (1995-2000); in FDA/ASA Biopharmaceutical Section Workshop Planning Committee (1998, 2000) He served in FDA CDER Biostatistics Career Path Committee (1999-2001) and in FDA CDER Reviewer Affairs Committee (1995-1997). **[STATEMENT]** I am deeply honored to be

nominated as the candidate of president of ICSA. My interest of running for the position is inspired by the efforts and contributions of my fellow members served for the association throughout the years. With their efforts, ICSA enjoyed all the successes since its inception. We all witnessed the rapid growth of the Association such that it is well recognized by our professional colleagues as an important international organization in the statistical profession. With the unselfish contributions and rigorous efforts of our members, ICSA provided many worthy services to its members. For examples, Statistica Sinica, a statistical journal co-sponsored by ICSA, is an influential and with high impact journal in our profession. The communicative ICSA Bulletin is one of the most popular and readable statistical newsletters. The ICSA Applied Statistics Symposium is one of the major statistical events in United States. The ICSA International Conference enjoyed its successes every four years. The book and journal donation program makes significant contributions in building up statistical libraries in China. However, with the growth of the association, there are more works to be done. First, ICSA has grown to the size that it is too large to be handled efficiently by members without professional business training and with only time and energy can be contributed during the weekends. It needs professional cares. Second, ICSA needs to enhance its web service capabilities such as membership renewal, improved online membership directory, credit card payment system, online statistical forum, real-time information delivery and communication, electronic election, etc. Efficient online service will not only improve communication and information delivery, it will also reduce the mailing cost that has been the major expense which is ever increasing. Third, ICSA needs a more applied statistical journal of its own. This journal will compliment and enhance the scope and importance of Statistica Sinica. Using internet as the medium, an publishing-as-it-goes e-journal can dispatch the author's finding on a more timely basis. Such a journal will increase the attraction to potential readers to ICSA. In order to achieve these goals, we will need more members and funding. If elected, I will focus on these priorities during my tour of duty and work hard with my fellow members to achieve them.

Candidates for 2005 Biometrics Section Chair

TAN, Ming T.

[PRESENT POSITION] Dr. Tan is Head, Division of Biostatistics, University of Maryland Greenebaum Cancer Center and Professor of Epidemiology and Preventive Medicine. He is also chief statistician for the University of Maryland Academic Public Private Partnership Program (funded by National Cancer Institute, NCI) to conduct novel cancer therapeutic, prevention, diagnostic, and imaging intervention-directed research. **[PREVIOUS POSITIONS]** He was Associate Member/Professor of Biostatistics at the Department of Biostatistics and Epidemiology, St. Jude Children's Research Hospital and director of Biostatistics for the St Jude Developmental Therapeutics for Solid Malignancies and Biostatistics Core Leader of the Solid Tumor Program Project funded by the National Cancer Institute. He was Assistant (1990-1996) and Associate Staff/Professor (1996-1997) of Biostatistics, Department of Biostatistics and Epidemiology at The Cleveland Clinic Foundation, where he collaborated in multiple therapeutic areas, most noticeably multi-center trials in cardiology, diabetes and renal diseases, sponsored by National Institute of Health (NIH) and industry. **[DEGREES]** Ph.D., 1990, Statistics, Purdue University, West Lafayette, Indiana, USA. MS in Mathematical Statistics from joint statistics program of Central China Normal University and National Wuhan University and BS from Wuhan University. **[FIELD OF MAJOR STATISTICAL ACTIVITIES]** Dr. Tan's research interests include decision theory, empirical Bayesian and dose-response models; frequentist approach to group sequential clinical trial design and the design and analysis of cancer drug development xenograft models and drug interactions funded by established investigator awards of NIH/NHLBI and NIH/NCI. His interests also include random-effects and Bayesian models for repeated measures data, the evaluation of biomarkers and/or diagnostic tests and innovative quantitative solutions to problems in oncological drug development and cancer research. **[PUBLICATIONS]** Include 34 peer-reviewed papers in statistical and biostatistical method journals such as *Biometrics, Statistica Sinica, Journal of Multivariate Analysis, Annals of the Institute of Statistical Mathematics, Statistics in Medicine,* and *Journal of Biopharmaceutical Statistics* and 40 medical papers in *New England Journal of Medicine, Proceedings of National Academy of Sciences,*

Circulation, Kidney International, Journal of Infectious Diseases, Cancer, Clinical Cancer Research, Cancer Chemotherapy and Pharmacology, and *Oncogene.* **[ICSA ACTIVITIES AND OFFICES HELD]** Member of ICSA and currently serves on ICSA's Membership and Awards Committees. Organizer and chair of Invited Paper Session in ICSA applied statistics symposium for three years. **[PROFESSIONAL ACTIVITIES]** Dr. Tan is currently associate editor of *Biometrics* and *Communications in Statistics: Theory and Methods.* He serves on multiple NIH and NCI review panels and is a member of FDA CDER advisory committee (1999-2003) and consultant for NDA endeavors of multiple pharmaceutical companies. He was organizer of invited paper session for International Biometrics Society ENAR Spring Meeting 1994. He is an elected member of the International Statistical Institute (1999). He is a member of the American Statistical Association, International Biometric Society, Society for Clinical Trials and Institute of Mathematical Statistics.

TING, Naitee

[PRESENT POSITION] Naitee Ting is currently an Associate Director of Biostatistics at Pfizer Global Research & Development in New London, CT. He has been with Pfizer Research since 1987. **[FORMER POSITIONS]** Naitee served a few different positions (including Programmer/Analyst, Demographer) in various organizations before obtaining his Ph.D. degree. **[DEGREES]** Ph.D. in Statistics, Colorado State University (1987); M.S. in Statistics, Mississippi State University (1979); B.S. in Forestry, College of Chinese Culture (Taipei, 1976). **[FIELDS OF MAJOR STATISTICAL ACTIVITES]** His major activities have been statistical applications in clinical research. Naitee has made contributions in constructing approximate confidence intervals on functions of variance components. His current interest is in the design and analysis of dose response studies.
[PUBLICATIONS] Naitee has published articles in variance components, drug safety, and clinical trials. His publications appear in Technometrics, Biometrical Journal, Drug Information Journal and others. Some of his articles are published as book chapters. **[ICSA ACTIVITIES AND OFFICES HELD]** Naitee was the Secretary for the Biometrics Section of ICSA between 1995 and 1996. He served as the Executive Director of ICSA for a 3-year term between 1998 and 2000. Naitee is currently a Board of Director of ICSA (2002-2004). **[RELATED PROFESSIONAL ACTIVITES]** He was Vice President (1990-1991) and President (1991-1992) of the Connecticut Chapter of ASA. Naitee served as the Council Representative of Connecticut Chapter (1997-1999). He was also a member of The Student Paper Competition Committee for the Biopharmaceutical Section (1997-1999), Naitee chaired this Committee in 1999.

Candidates for Board of Directors – 2005 to 2007

HSIAO, Chin-Fu

[Present Position] Assistant Investigator, Division of Biostatistics and Bioinformatics, National Health Research Institutes, Taiwan. **[Former Position]** Postdoctor Fellow, Division of Biostatistics and Bioinformatics, National Health Research Institutes, Taiwan.
[Degrees] Ph.D, 1997, Statistics, University of Wisconsin-Madison. M.S., 1989, Mathematics, National Central University. B.S., 1987, Mathematics, National Central University, Taiwan. **[Fields of Major Statistical Activities]** Sequential decision theory; Bayesian analysis; genetic study. **[Selected Publications]** Liu, JP, Hsueh, HY and Hsiao, CF. (2004). Bayesian non-inferiority approach to evaluation of bridging studies. To appear in Journal of Biopharmaceutical Statistics. Hsiao, CF, Xu, JZ and Liu JP. (2003). Group sequential approach to evaluation of bridging studies. Journal of Biopharmaceutical Statistics 13(4): 793-801. Liu JP, Hsiao, CF and Hsueh HY. (2002). Bayesian approach to evaluation of bridging studies. Journal of Biopharmaceutical Statistics 12(3):401-408. Ko, J-H, Lee, T-C, Hsiao, CF, Lin, G-L, Liang, C-W, Chen, T-T, Wang, L-D, Chao, C-C, Yen, S-H, Chen, K-Y, Sheen, T-S, Hsiung, C-A, Chen, P-J, Hsu, M-M, and Jou, Y-S (2002). A comprehensive deletion mapping on chromosomes 3, 9, 11 and mutational screening of FHIT, $p16^{INK4a}$, $p19^{ARF}$ genes in nasopharyngeal carcinoma. Cancer. 94(7): 1986-1996. Wu, K-D, Hsiao, CF, Ho, L-T, Sheu, H-H, Pei, Dee, Curb, D., Chen, Y-D I., Tsai, H-J, Dzau, V. J., Cox, D., Tai, T-Y (2002). Clustering and heritability of insulin resistance in Chinese and Japanese hypertensive families: A SAPPHIRe sib study. Hypertension Research, 25, 529-536. Chang, I. S., Chen, M. Y., Hsiao, CF, and Hsiung, C. A. (2002). A unified multipoint linkage analysis of qualitative and quantitative traits for sib-pairs. Statistica Sinica 12, 297-

309. Hsiao, CF and Clayton, Murray K. (2001). Bayes discrete sequential boundaries for clinical trials. Communications in Statistics: Theory and Method 30, 1381-1394. Hsiao, CF and Clayton, Murray K. (2001). Lerche's sequential test for the drift of a Brownian motion with a smooth prior. Sequential Analysis 20, 183-189. Ranade, K., Wu, K.D., Risch, N., Olivier, M., Pei, D., Hsiao, CF, Chuang, L.M., Ho, L.T., Jorgenson, E., Pesich, R., Chen. Y.D.I., Dzau, V., Lin, A., Olshen, R.A., Curb, D., Cox, D.R., and Botstein D. (2001) Genetic Variation in aldosterone synthase predicts plasma glucose levels, PNAS 98, 13219-13224. Ranade, K., Chang, M. S., Ting, C. T., Pei, D., Hsiao, CF, Pesich, R., Hebert, J., Chen, Y. D., Olshen, R., Risch, N., Cox and D. R., Botstein, D. (2001). High-throughput genotyping and mapping of single nucleotide polymorphisms. Genome Research 11, 1269-1274. **[ICSA Activities and Offices Held]** Member of ICSA, participated in past symposium. **[Related Professional Activities]** Organizer of "2003 Symposium on Statistical Methodology for Evaluation of Bridging Evidence," Taipei.

HUANG, Jian

[PRESENT POSITION] Associate Professor, Department of Statistics and Actuarial Science, and Program in Public Health Genetics, University of Iowa. **[FORMER POSITION]** Asistant Professor, Department of Statistics and Actuarial Science, University of Iowa, 08/1994-04/1999. **[DEGREES]** Ph.D.,1994, University of Washington. M.S., 1987, Wuhan University. B.S., 1985, Wuhan University. **[FIELDS OF MAJOR STATISTICAL ACTIVITIES]** Semiparametric Inference, Statistical Genetics, and Survival Analysis **[SELECTED PUBLICATIONS]** Huang, J. and Jiang Y. (2003). Genetic Linkage Analysis of a Dichotomous Trait Incorporating a Tightly Linked Quantitative Trait in Affected Sib Pairs. Am J Hum Genet 72: 949-960. Wang, K. and Huang, J. (2002). A Score Statistic Approach for Mapping Quantitative Trait Loci with Sibships of Arbitrary Size. Am J Hum Genet, 70: 412-424. Huang, J., Vieland V.J. and Wang, K. (2001). Nonparametric Estimation of Marginal Distributions under Bivariate Truncation with Application to Testing for Age-of-onset Anticipation. Statistica Sinica, 11, 1047-1068. Slager, S. L., Huang, J. and Vieland, V. J. (2000). The Effect of Allelic Heterogeneity on the Power of Transmission Disequilibrium Test. Genet Epi, 18: 143-156. Huang, J. (1999). Efficient Estimation of the Partly Linear Additive Cox Model. Ann Statist, 27, 1536-1563. Huang, J. and Wellner, J. A. (1997). Interval Censored Survival Data: A Review of Recent Progress. Proceedings of the First Seattle Symposium in Biostatistics: Survival Analysis. Eds. D. Lin and T. Fleming. Springer-Verlag, New York. Huang, J. (1996). Efficient Estimation for the Proportional Hazards Model with Interval Censoring. Ann Statist 24, 540-568. **[ICSA ACTIVITIES AND OFFICES HELD]** Member of ICSA. **[RELATED PROFESSIONAL ACTIVITIES]** Vice president of the Iowa Chapter of the ASA(1998). President of the Iowa Chapter of the ASA (1999). Grant review for NIH, NSF, NSA, and NSERC of Canada. NSF Statistics and Probability Screening Panel (2001). Referee for statistics and genetics journals.

LEE, J. Jack

[PRESENT POSITION] Professor, Department of Biostatistics and Applied Mathematics, University of Texas M.D. Anderson Cancer Center, since 2001. **[FORMER POSITIONS]** Associate Professor (1997-2001); Assistant Professor (1991-1997), University of Texas M.D. Anderson Cancer Center. Adjunct Assistant Professor (1989-1991), Department of Biostatistics, University of California, Los Angeles. Stanley Schor Visiting Scholar, Merck Research Laboratories (1998-1999). **[DEGREES]** Ph.D., 1989, Biostatistics, M.S., 1984. University of California, Los Angeles. D.D.S., 1982, National Taiwan University. **[FIELDS OF MAJOR STATISTICAL ACTIVITIES]** Clinical Trial Design and Analysis; Biomarkers; Longitudinal Data; Survival Analysis; Statistical Graphics. **[PUBLICATIONS]** Published over 130 articles in mainstream statistical and medical journals, including Journal of the American Statistical Association, Biometrics, Journal of Computational and Graphical Statistics, Controlled Clinical Trials, The American Statistician, Statistics in Medicine, New England Journal of Medicine, Nature Medicine, Journal of the National Cancer Institute, Journal of Clinical Oncology, Cancer Research, Clinical Cancer Research, Cancer Epidemiology, Biomarker, and Prevention, Cancer, etc. **[ICSA ACTIVITIES]** Permanent Member of ICSA, participated in several ICSA Applied Statistics Symposia. **[RELATED PROFESSIONAL ACTIVITIES]** Statistical Editor, Journal of the National Cancer Institute, 2001-present, Received research grants from NIH, and DOD. Participated in several NIH Special Review Panels, Referee for statistical and medical journals. Served in IRB and Data Safety and Monitoring Board of several clinical trials. Chair-Elect of Faculty Senate, University of Texas M.D. Anderson Cancer Center; Member of Faculty Advisory Council, the University of Texas System.

LIU, Aiyi

[PRESENT POSITION] Investigator, Biometry and Mathematical Statistics Branch, Division of Epidemiology, Statistics and Prevention Research, National Institute of Child Health and Human Development, National Institutes of Health, Department of Health and Human Services, Bethesda, MD. **[FORMER POSITION]** Assistant Professor (1999-2002), Georgetown University Medical Center, Washington, DC; Postdoctoral Research Fellow (1997-1999), St. Jude Children's Research Hospital, Memphis, TN. **[DEGREE]** Ph.D in Statistics and Biostatistics, University of Rochester, Rochester, NY, 1997; M.S. in Statistics, University of Rochester, Rochester, NY, 1995; M.S. in Statistics, University of Science and Technology of China, 1988; B.S. in Mathematics and Statistics, University of Science and Technology of China, 1986. **[FIELDS OF MAJOR STATISTICAL ACTIVITIES]** Estimation theory; Sequential methodology in clinical trials; Regression analysis and multivariate data analysis; Receiver operating characteristics curves analysis and diagnostic medicine. **[PUBLICATIONS]** (Selected from over 40 papers published in peer-reviewed journals) **Liu A**, Schisterman EF, Zhu Y. On linear combinations of biomarkers to improve diagnostic accuracy. *Statistics in Medicine*. In Press. **Liu A**, Schisterman EF, Teoh E. (2004). Sample size and power calculation in comparing diagnostic accuracy of biomarkers with pooled assessments. *Journal of Applied Statistics* 31: 49-59; **Liu A**, Schisterman EF (2004). Principal component analysis. In *Encyclopedia of Biopharmaceutical Statistics*, 2nd ed., Chow SC, Editor, 1-6; Blancato J, Singh B, **Liu A**, Liao DJ, Dickson RB (2004). Correlation of amplification and over-expression of the c-*myc* oncogene in high-grade breast cancer: FISH, *in situ* hybridisation, and immunohistochemical analyses. *British Journal of Cancer* 90:1612-1619; **Liu A** (2003). A simple low-bias estimate following a sequential test with linear boundaries. In: *Crossing Boundaries: Statistical Essays in Honor of Jack Hall*, IMS Monograph Series, Kolassa J and Oakes D, Editors, 47-58; **Liu A**, Schisterman EF (2003). Comparison of diagnostic accuracy of biomarkers with pooled assessments. *Biometrical Journal* 45: 631-644; Mazumdar M, **Liu A** (2003). Group sequential design for comparative diagnostic accuracy studies. *Statistics in Medicine* 22: 727-739; **Liu A**, Shih WJ, Gehan E (2002). Sample size and power determination for clustered repeated measurements. *Statistics in Medicine* 21: 1787-1801; **Liu A** (2002). Efficient estimation of two seemingly unrelated regression equations. *Journal of Multivariate Analysis* 82: 445-456; **Liu A**, Zhang Y, Gehan E, Clarke R (2002). Block principal component analysis with application to gene microarray data classification. *Statistics in Medicine* 21: 3465-3474; Hall WJ, **Liu A** (2002). Sequential tests and estimates after overrunning based on maximum-likelihood ordering. *Biometrika* 89: 699-707; Wong LJ, Wong IH, **Liu A** (2002). Intergenerational transmission of pathogenic heteroplasmic mitochondrial DNA. *Genetics in Medicine* 4: 78-83; Poola I, Abraham J, **Liu A** (2002). Estrogen receptor beta splice variant mRNAs are differentially altered during breast carcinogenesis. *Journal of Steroid Biochemistry and Molecular Biology* 82: 169-179; **Liu A**, Hall WJ (2001). Unbiased estimation of secondary parameters following a sequential test. *Biometrika* 88: 895-900; **Liu A**, Boyett J, Xiong XP (2000). Sample size calculation for planning group sequential longitudinal trials. *Statistics in Medicine* 19: 205-220; **Liu A**, Tan M, Boyett J, Xiong XP (2000). Testing secondary hypotheses following sequential clinical trials. *Biometrics* 56: 640-644; **Liu A** (2000). Maximum likelihood estimate following sequential probability ratio tests. *Sequential Analysis* 19: 63-75; **Liu A**, Hall WJ (1999). Unbiased estimation following a group sequential test. *Biometrika* 86: 71-78; **Liu A**, Hall WJ (1999). Minimum variance unbiased estimation of the drift of Brownian motion with linear stopping boundaries. *Sequential Analysis* 17: 91-107, 1998; **Liu A** (1997). On the maximum likelihood estimate for the drift of Brownian motion following a symmetric sequential probability ratio test. *Communications in Statistics---Theory and Methods* 26: 977-989; **Liu A** (1994). Selection of covariates and estimation of parameters in growth curve models. *Acta Mathematica Sinica* 37: 362-372; **Liu A** (1993). An efficient estimation of seemingly unrelated multivariate regression models with application to growth curves analysis. *Statistica Sinica* 3: 421-434. **[ICSA ACTIVITIES]** Permanent member of ICSA; Member of Planning Committee of 2005 Applied Statistics Symposium. **[RELATED PROFESSIONAL ACTIVITIES]** Active member of ASA and ENAR.

LIU, Guanghan (Frank)

[PRESENT POSITION] Associate Director, Clinical Biostatistics, Merck Research Laboratories, West Point, PA **[FORMER POSITION]** Biometrician and Sr. Biometrician (1995-2001), Clinical Biostatistics, Merck Research Laboratories, West Point, PA, Postdoctoral Research Fellow (1994-1995), Department of Biostatistics, The Johns Hopkins University, Baltimore, MD; Research Statistician (1993-1994), Graduate School of Education, University of California, Los Angeles, CA **[DEGREE]** Ph.D in Statistics, University

of California, Los Angeles, CA, 1993; M.S. in Mathematics, University of California, Los Angeles, CA, 1990; M.S. in Statistics, East China Normal University, China, 1987; B.S. in Mathematics, East China Normal University, China, 1984.

[FIELDS OF MAJOR STATISTICAL ACTIVITIES] Clinical trial design and data analysis; sample size and power analysis for clinical studies; longitudinal data analysis; missing data methodology; analysis of survival data. **[PUBLICATIONS]** **Liu, G.**, and Gould, AL (2002), "Comparison of alternative strategies for analysis of longitudinal trials with dropouts", Journal of Biopharmaceutical Statistics 12: 207-226; **Liu, G.** (2000), Sample Size for Epidemiologic Studies, *Encyclopedia of Epidemiologic Methods*, Edited by Mitchell Gail and Jacques Benichou, John Wiley & Sons; **Liu, G.** (1999), An application of multiple imputation in longitudinal clinical studies with dropouts. *Proceedings of the Biopharmaceutical Section*, ASA, 84-89; **Liu, G.**, Jiang, K., Getson, A., Gould, L. and Polis, A. (1997), Analysis of grouped survival data in cross-over trials using GEE approach. *Proceedings of the Biopharmaceutical Section*, ASA, 272-277; **Liu, G.** and Liang, K.Y.(1997), Sample size and power calculations for studies with correlated observations, *Biometrics*, 53: 937-947; Piantadosi, S. and **Liu, G.** (1996). Improved designs for dose escalation studies using pharmacokinetic measurements, *Statistics in Medicine*, 15: 1605-18; **Liu, G.** (1994). A locally correlated process and its applications to Bayesian estimation. *Journal of Multivariate Analysis*, 49, No. 1, 132-149; **Liu, G.** (1989). Equivalent marginal multivariate exponential distribution and its parameter estimation (Chinese). *Chinese Journal of Applied Probability and Statistics*, 5, 295-302. **[ICSA ACTIVITIES]** Member of ICSA since 1995; Attendee and presenter of ICSA Applied Statistics Symposiums and ICSA International Conference. **[RELATED PROFESSIONAL ACTIVITIES]** Active member of ASA and ENAR; Editor for BioPharm Quarterly, an on-line journal for Division of Bio/Pharmaceutical Sciences, Society of Chinese Bioscientists in America; Referee for statistical journals including Biometrics, Statistics in Medicine, Journal of Biopharmaceutical Statistics, The American Statistician, and Controlled Clinical Trials.

LU, Ying

[PRESENT POSITION] Associate Professor of Radiology and Biostatistics, Department of Radiology/Epidemiology and Biostatistics; Director of Biostatistics Core, UCSF Comprehensive Cancer Center, University of California, San Francisco. **[FORMER POSITION]** Assistant Professor of Radiology, UCSF (1994-1998); Assistant Professor of Epidemiology, University of Miami School of Medicine, FL (1990-1994); Assistant Teacher, Shanghai Jiao Tong University (1984-1985). **[DEGREE]** Ph.D., Biostatistics, 1990, University of California, Berkeley; MS, Applied Mathematics, 1984, Shanghai Jiao Tong University; BS, 1982, Mathematics, Fudan University. **[HONORS]** The Evelyn Fix Memorial Award, 1990; Healthstar Osteoporosis Medical-Research Award, 2003. **[FIELD OF MAJOR STATISTICAL ACTIVITY]** Research, Teaching, and Consulting in various statistical areas: Statistical methods for medical diagnosis; cancer clinical trials; Cost-effective diagnostic tests; Osteoporosis and Arthritis; Quality control and quality assurance; meta-analysis; stochastic process for chronic diseases; and statistical methods for animal carcinogenicity experiments. **[PUBLICATIONS]** To the end of April 2004, he has 104 papers appeared in peer-reviewed journals (Statistics in Medicine, Medical Decision Making, Radiology, Journal of Bone and Mineral Research, Osteoporosis International, Biometrics, Mathematical Biosciences, Cancer, International Journal of Epidemiology, American Journal of Epidemiology, Stroke, NeuroImaging, etc.); 8 in press; 5 comprehensive reviews and letters; 11 book chapters; and co-editor of one book. Recent representative statistical publications include: Jin, Lu, et al, Classification algorithm for hip fracture prediction based on recursive partitioning methods. *Medical Decision Making* (in press); Lu, Jin, Mi, On comparison of two classification methods with survival endpoints. *Handbook of Statistics*, Vol. 23:43-59, 2004; Lu, Jin, Genant, On the equivalence of two diagnostic tests based on paired observations. *SIM* 2003 Oct; 22(10):3029-44; Lu, Zhao, Heller, ROC analysis for diagnostic examinations with uninterpretable cases, *SIM* 2002, 21:1849-1865; **[ICSA ACTIVITIES]** Member of Program Committee, Chair of Short-Course subcommittee ICSA 2004 Applied Statistics Symposium. Local coordinator for ICSA activities in the 2003 JSM **[RELATED PROFESSIONAL ACTIVITIES]** He was the president (1998-1999) and vice president for Biostatistics (1995-97) of SF Bay Area ASA Chapter; Member of International Committee for Standards in Bone Measurement (1995-2000). He also organized invited sessions for JSM (1997) and American Public Health Association Annual Meetings (1995, 1996).

SUN, Jiayang

[PRESENT POSITION] Professor, Dept. of Statistics, Case Western Reserve University. **[FORMER POSITIONS]** Associate Professor, Department of Statistics, Case Western Reserve University; Assistant Professor, University of Michigan. **[DEGREES]** Ph.D., 1989, Statistics, Stanford University; M.S., 1985, Peking University; B.S., 1982, Mathematics, Anhui University. **[FIELDS OF MAJOR STATISTICAL ACTIVITIES]** Biased sampling, bump hunting, images and mixtures; Semiparametrics and nonparametrics; Simultaneous Inferences; Statistical computing, data mining and high dimensional graphics; Statistics in Astronomy, IT and medical sciences. **[SELECTED PUBLICATIONS]** Dr. Sun published in a wide variety of journals. Representative publications: Significance Levels in Exploratory Projection Pursuit, *Biometrika*, and (1991) Vol. 78, No. 4, pp. 759-769. A Penalized Likelihood Estimate of When Is Non-increasing, Statistica *Sinica*, (1993) Vol. 3, No. 2, pp. 501-515. Some Practical Aspects of Exploratory Projection Pursuit, *SIAM J. Sci. Stat. Comp.*, (1993) Vol. 14, No. 1, pp. 68-80. Tail Probabilities of the Maxima of Gaussian Random Fields, *Annals of Probability*, (1993) Vol. 21, No. 1, pp. 34-71. Sums and Maxima of Discrete Stationary Processes, Journal *of Applied Probability*, (1993) Vol. 30, pp. 863-876. Simultaneous Confidence Bands for Linear Regression and Smoothing, *Annals of Statistics*, (1994) Vol. 22, No. 3, pp. 1328-1346. Simultaneous Confidence Bands for Linear Regression with Heteroscedastic Errors, *JASA*, (1995) Vol. 90, No 431, 1094-1098. Adaptive Smoothing for a Penalized NPMLE of a Non-increasing Density, *JSPI*, (1996) Vol 52, pp 143-159. Robustness of Tube Formula, *JCGS*, (1997) Vol 6, No 2, pp242-50. Semi-parametric Estimation for Biased Sampling Models," *Statistica Sinica*, (1997) Vol 7, No 3, 545-576. Tests in Projection Pursuit Regression, *JSPI*, (1998) Vol 75, No 1, 65-90. Testing Uniformity Versus A Monotone Density, Annals of Statistics, (1999) Vol 27, and No 1, 338-360. Simultaneous Confidence Bands for Growth and Response Curves, *Statistica Sinica*, (1999) Vol. 9, No. 3, 679-698. Confidence Bands in Generalized Linear Models, Annals *of Statistics*, (2000) Vol 28, No. 2, 429-460. Bayesian Model Selection in Finite Mixtures by Marginal Density Decompositions, *JASA*, (2001) Vol. 96, No. 456, 1316-1332. Multiple Comparisons for a Large Number of Parameters. *Biometrical Journal*, (2001) Vol 43, No 5, 627-643. Automated Support for Classifying and Prioritizing Software Failure Reports, Intl. Conf. On Software Engineering, (2003). Testing Homogeneity in a Mixture Distribution via the Distance Between Competing Models, (2004),Vol.99, No 466, pp. 488-498.**[ICSA ACTIVITIES]** Permanent Member of ICSA and organized an invited session for the ICSA 2000 Applied Statistics Conference. **[RELATED PROFESSIONAL ACTIVITIES]** Elected member of ISI. Research funded by NSF since 1990 and by NIH since 2003. Organizer of the International Workshop on Developments and Challenges in Mixture models, Bump Hunting and Measurement Error Models. IMS program chair of 2002 IMS/ENAR meetings. IMS treasurer-elect (2004-2007) and exec member (2004-2007). Member of ASA committee on SRCOS (2004-2006), and ASA Committee on Minority in Statistics (1995-1999).

WANG, Naisyin

[PRESENT POSITION] Professor, Department of Statistics and Toxicology, Texas A&M University, College Station, TX**[FORMER POSITION]** Assistant and Associate Professor (1992-2001) Department of Statistics, Texas A&M University, College Station, TX. **[DEGREE]** Ph.D in Statistics, Cornell, 1992; A.M.S. in Applied Statistics, Ohio State University, 1987; B.S. in Mathematics, Tsinghua University, Taiwan, 1986. **[FIELDS OF MAJOR STATISTICAL ACTIVITIES]** Semiparametric and nonparametric models. Measurement error problems; Imputation estimation; Longitudinal data analysis; Nonlinear mixed models; Transformation and weighting; Mixture modeling. **[PUBLICATIONS]** Professor Wang has published over 50 papers in statistical and biological science journals, including JASA, Biometrika, JRSSB, Statistica Sinica, Biometrics, Biostatistics, Journal of Statistical Planning and Inference, Communications in Statistics, Journal of Medicinal Chemistry, Journal of Toxicology and Environmental Health, Cell Growth and Differentiation, Carcinogenesis, Cancer Epidemiology, Biomarkers & Prevention, American Journal of Veterinary Research, Veterinary Therapeutics, Advanced Experimental Medical Biology, Journal of Biomedical Optics, Journal of Lipid Research. **[PROFESSIONAL SERVICES]** Permanent member, NIH SNEM-5 Study Section (2004-2008); Panel member, NSF grants (2001); Chair, ENAR Young Researcher Workshop, 2002; Member, COPSS Fisher Lecture Selection Committee, 2003-2006; Chair, WNAR/IMS meeting IMS program committee, 2003; Member, ENAR Regional Advisory Board, 2000-2003; Member, ENAR Student Award Committee, 2003-2004; Associate Editor, Biometrics (1997-2004); Associate Editor, Communication in Statistics (1998-2001); Associate Editor, JASA (1999-2003) **[ICSA ACTIVITIES]** Permanent Member of ICSA; Invited Session Organizer, ICSA Applied Statistics Symposium, 2000; Member, Public Relations Committee, 1999-2001.

YANG, Song

[PRESENT POSITION] Senior Mathematical Statistician, Office of Biostatistics Research, National Heart, Lung, and Blood Institute, since 2003. **[FORMER POSITIONS]** Statistics Coordinator, 2001-2003, Professor, 2000-2003, Associate Professor with tenure, 1995-2000, Assistant Professor, 1990-1995, Visiting Assistant Professor, 1988-1990, Department of Mathematics and Statistics, Texas Tech University. Visiting Associate Member, Department of Biostatistics, Fred Hutchinson Cancer Research Center. 1997-1998. **[DEGREE]** Ph.D., 1988, Statistics, Michigan State University. B. S., Mathematics, 1982, Sichuan University. **[FIELDS OF MAJOR STATISTICAL ACTIVITIES]** Survival Analysis, Reliability, Genetic Epidemiology, Clinical Trials, Analysis of Clustered Data, Semiparametric Regression, Robust Inference, and Nonparametrics. **[SELECTED PUBLICATIONS]** Yang and Prentice (2004), Semiparametric analysis of short-term and long-term hazard ratios with two-sample survival data. To appear in Biometrika; Yand, Hsu and Zhao (2004). Combining asymptotically normal tests: cases studies in comparison of two groups. To appear in J. of Statistical Planning and Inference; Hummer and Yang (2001). On the extended log-rank estimating function for the censored regression model. Computational Statistics and Data Analysis. 37, 171-180; Yang (2000). Functional estimation under interval censoring case 1. J. of Statistical Planning and Inference 89, 135-144; Regalado Yang and Wesson (2000). Prospective analysis of the effect of cigarette smoking on progression of renal insufficiency in severe essential hypertension.
American Journal of Kidney Diseases 35, 687-694; Yang and Prentice (1999). Semiparametric inference in the proportional odds regression model. J. Amer. Statist. Assoc. 94, 125-136; Yang (1999). Censored median regression using weighted empirical hazard and survival functions. J. Amer. Statist. Assoc. 94, 137-145; Yang (1998). Some scale estimators and lack-of-fit tests for the censored two sample accelerated life model. Biometrics 54, 1040-1052; Yang (1997). A generalization of the product-limit estimator with an application to censored regression. Ann. Statist. 25, 1088-1108; Yang (1997). Extended weighted log-rank estimating functions in censored regression. J. Amer. Statist. Assoc. 92, 977-984; Ruymgaart and Yang (1997). Some applications of Watson's perturbation approach to random matrices. J. Multivar. Analysis 60, 48-60. Yang (1996). Weighted empiricals and the product-limit estimator in the multiplicative hazard and time transfer regression model. Statist. and Probab. Lett. 30, 17-24; Yang (1996). One-step robust estimation in parametric models. Statist. and Probab. Lett. 26, 225-232; Yang (1995). Koul-Dewit type estimation in censored regression using the conditional empirical process. J. of Statistical Planning and Inference 46, 351-363; Yang and Koul (1995). Minimum distance estimation of the center of symmetry with randomly censored data. Metrika 42, 79-97; Yang (1993). A central limit theorem of the integrated square error of the kernel density estimators with randomly censoreddata. J. of Statist. Planning and Inference, 37, 127-143; Yang (1992). Some inequalities about the Kaplan-Meier estimator. Ann. Statist. 20, 535-544; Yang (1991). Minimum Hellinger distance estimation of parameters in the random censoring model. Ann. Statist. 19, 579-602. Karunamuni and Yang (1991). Weak and stronguniform consistency rates of kernel density estimates for randomly censored data. Canad. J. of Statist. 19, 349-359; Yang (1990). Efficient robust estimation of parameter in the random censoring model. Statist. and Probab. Lett. 10, 419-426. **[RELATED PROFESSIONAL ACTIVITIES]** Member of ASA and IMS; Adjunct Professor, Sichuan University; Associate Editor, Statist. and Probab. Lett.; Referee for jounals including Ann. Probab., Ann. Statist., J. Amer. Statist. Assoc., J. Multivar. Statisit., J. Royal Statist. Soc. (B), J. Statisit. Planning and Infer., LifeTime Data Analysis., Scand. J. Statisit., Statist. and Probab. Lett., and Statistica Sinica; Reviewer for NSF and NIH proposals, and NIH protocols.

ZHU, Guangrui Ray

[PRESENT POSITION] Director Global Biostatistics, Aventis Pharmaceuticals, Bridgewater, NJ. **[FORMER POSITIONS]** Statisticians at Pfizer, Schering-Plough, and US FDA since 1990. **[DEGREES]** Ph.D. in Statistics, 1992 Stanford University. **[FIELDS OF MAJOR STATISTICAL ACTIVITIES]** Clinical trial methodology: self-tuning clinical trial design, non-inferiority studies; Pharmacokinetics and Pharmacodynamics modeling: non-linear mixed effect models and population PK/PD; Adaptive non-linear estimation and prediction; Using bootstrap and non-linear mixed effect modeling to establish bio-equivalence based on PD markers for metered dose inhalers. **[SELECTED PUBLICATIONS]** "Statistics in Pharmacology and Pre-clinical Studies", a chapter in Advanced Medical Statistics, 2001, with T.L. Lai and M.C. Shih; "Predictors of an acute antidepressant response to fluoxetine and sertraline", International Clinical Psychopharmacology, 1999, with M.F. Flament, R. Lane and Z. Ying; "The effect of age on the clearance of felbamate: a retrospective analysis using non-linear mixed effects modeling", Therapeutic Drug Monitoring, 1997, with C. Banfield and P. Glue;

"Adaptive Prediction in Non-linear Autoregressive Models and Control Systems", Statistica Sinica, 1991, with T.L. Lai. **[ICSA ACTIVITIES AND OFFICES HELD]** Organized and chaired a session of Population Pharmaco-Statistic Modeling in ICSA 1996 Applied Statistics Symposium. **[RELATED PROFESSIONAL ACTIVITIES]** Program committee member for a Workshop in Beijing, 2003 on "Biostatistics in Clinical Research for Drug Development and Evaluation"; DIA and ASA activities.

Zhu, Lixing

[PRESENT POSITION] Associate Professor, Department of Statistics and Actuarial Science, The University of Hong Kong and the Adjunct Professor of Statistics, East China Normal University, Shanghai, China **[FORMER POSITIONS]** Professor, Academy of Mathematics and System Science, Chinese Academy of Sciences **[DEGREES]** PHD in Probability and Statistics Chinese Academy of Sciences (1990), MS in Statistics, University of Science and Technology of China (1985) and BS in Mathematics, Anhui University (1982). **[FIELDS OF MAJOR STATISTICAL ACTIVITIES]** Large scale dada analysis, Nonparametric Monte Carlo tests, Regression analysis, Survival analysis, Empirical process theory. **[SELECTED PUBLICATIONS]** Dr. Zhu has published 96 papers including Zhu, Lixing and Fang Kai Tai (1996) Asymptotics for kernel estimates of sliced inverse regression, *Ann. Statist.* **24**, 1053-1068; Stute, W., Thies, S. and Zhu, L. X. (1998) Model checks for regression: An innovation process approach, *Ann. Statist.*, **26**, 1916-1934; Zhu, L. X. and Neuhaus, G. (2000) Nonparametric Monte Carlo tests for multivariate distributions. *Biometrika,* 87, 919-928; Xia, Y., Li, W. K. and Tong, H., Zhu, L. X. (2002). An adaptive estimation of optimal regression subspace, *Journal of Royal Statistical Society, Series B*, **64**, 363-410. Read Paper (it was read in February meeting organized by Royal Statistical Society of Kingdom); He, X. and Zhu, L. X (2003). A Lack-of-Fit Test for Quantile Regression, *Journal of the American Statistical Association*, **464**, 1013-1022. **[HONORS]** *Fellow* of Institute of Mathematical Statistics; *Elected member* of International Statistical Institute; *Humboldt Research Award* granted by Alexander-von Humboldt Foundation of Germany in 2000. He is the *first awardee* in Natural Science and Engineering from Taiwan, Hong Kong and the mainland of China and the *only winner* in Statistics from Asia. **[ICSA ACTIVITIES AND OFFICES HELD]** The member of the organizing committee of the 4th and 5th ICSA conference. **[RELATED PROFESSIONAL ACTIVITIES]** a standing member of the council of Statistics and Probability Society of China (1994--2002), a standing member of the editorial board for "Journal of Applied Probability and Statistics", an Associate Editor of "Acta Mathematica Applica Sinica" and Editor of "International Journal of Systems Science and Complexity".

Congratulations!

Professor C. F. Jeff Wu, Coca Chair in Engineering Statistics, School of Industrial and Systems Engineering, Georgia Institute of Technology, was elected to the National Academy of Engineering (NAE) in 2004. His citation reads as "For conceiving and building modern systems of experimental design based on contemporary methods for parameter estimation to provide quality improvements."

Election to the National Academy of Engineering is among the highest professional distinctions accorded to an engineer. Academy membership honors those who have made "important contributions to engineering theory and practice, including significant contributions to the literature of engineering theory and practice," and those who have demonstrated accomplishment in "the pioneering of new fields of engineering, making major advancements in traditional fields of engineering, or developing /implementing innovative approaches to engineering education."

Wu is one of few statisticians ever elected to the NAE and arguably the first mathematical statistician to be on the membership.

Interview with a Distinguished Statistician

A Conversation with Dr. Jon Kettenring

By: Kao-Tai Tsai, Ph.D.

Dr. Jon Kettenring recently completed nearly 35 years in industrial research at Bell Labs, Bellcore, and Telcordia Technologies. Throughout his career he has devoted much of his energy to supporting organizations that are the backbone of the statistics profession, including service as President of the American Statistical Association (ASA) in 1997 and Chair of the Board of Trustees of the National Institute of Statistical Sciences (NISS) from 2000 to 2004. He recently became a Fellow in a program for retired industrial scientists at Drew University in Madison, NJ.

I would like to thank you first for taking time for me to conduct this interview. You have a distinguished career as a statistician. Would you please tell us a little bit of your personal history? And how did you get interested at statistics?

I went to Stanford University as an undergraduate student thinking that I would become an engineer. I tried several different branches, but none seemed quite right for me. I did not really decide my major field until my senior year at Stanford. After my junior year, I had an opportunity to participate in an overseas study program in Germany. I had a great time over there. We only needed to go to school four days a week, leaving plenty of time for travel and getting to know the local people. After I came back to Stanford, I concentrated on courses in mathematics and statistics and finally graduated with a statistics major. I also received an R.O.T.C. commission as a Second Lieutenant in the Army but decided to complete a masters degree in statistics before fulfilling my service obligation. I especially remember taking multivariate analysis from Professor Ingram Olkin. His lectures were dazzling and I was fascinated by the subject matter. I talked with him one day about continuing my studies after my military service, and he suggested that I consider the University of North Carolina. I did eventually follow his advice after I returned from 18 months in Turkey. This was in the fall of 1964. Perhaps the best thing that happened to me in Chapel Hill was the opportunity to work with Professor Norman Johnson, who was my thesis advisor. I especially appreciated his broad perspective on statistics, ranging from the theoretical to the practical and computational aspects.

Right to left: Dr. Kettenring, Professor Olkin, Dr. Karr.

You were elected the President of ASA previously. I still remember voting for you. What was your experience of holding the highest office of the largest statistical association?

It was a great honor and very satisfying to be the President of the ASA. ASA is the

President of the ASA. ASA is the largest statistical association in the world with a substantial international membership. The experience broadened my appreciation for the rich variety of contributions that we as statisticians make to society. More than anything else I enjoyed meeting hundreds of people that I had never met before. I had many opportunities to visit local ASA chapters across the country, and this was particularly stimulating for me.

There were other very special occasions, too. My wife and I had the pleasure of having dinner in the State Department as part of a ceremony honoring U.S. participants in the Mathematics Olympiad. I think that one evening fully justified in my wife's eyes all the years that I had spent working on ASA activities. She marveled at the furniture and the art work! The following year, when David Moore was ASA president, I had the opportunity to meet the late Senator Daniel Patrick Moynihan at the JSM. We were so lucky to have leadership of that caliber in the U.S. Senate, and I was proud to have had the chance to interact with him.

As the President, I also had the task to write monthly columns in the Amstat News. I was quite surprised to discover that many members actually read them very carefully, and some even wrote back to me about their thoughts of my columns. I still get occasional comments from people recalling that I said this or that, which is very gratifying.

What is the level of interactions between ASA and other statistical associations besides the joint sponsorship of conferences? I have been a member of ASA since I finished my graduate work and I have not seen much of anything beyond this.

ASA has many interactions with other statistical associations. For example, during the planning stage of the Joint Statistical Meetings, all the contributing associations need to have close interactions to plan the various sessions and other activities. Another example is the Committee of Presidents of Statistical Societies, known simply as COPSS. It sponsors many important awards and works to improve communications and coordination across different statistics communities.

What kinds of interactions does ASA have with other non-statistical disciplines? Ideally, what would you like to see?

Good question! We talk about this quite a bit but it always seems to be an uphill battle. Thanks to the Interface meetings and efforts of the Computing and Graphics sections of ASA, among others, we now have much stronger ties with computer science than we used to, which is crucial to our future.

At the moment I and several others from the statistics and operations research communities are investigating how we can advance science by bringing our two fields closer together. ASA and INFORMS held an NSF-sponsored workshop in Santa Fe last January to begin this process, and a summary report of the outcomes is being prepared. Fortunately, ASA and INFORMS already have close ties. I am optimistic that important actions will follow.

Interdisciplinary activity has really blossomed over the last decade. It is the basis of the entire research program at NISS. And it is natural for ASA to continue to provide leadership that will support and encourage these developments across the association.

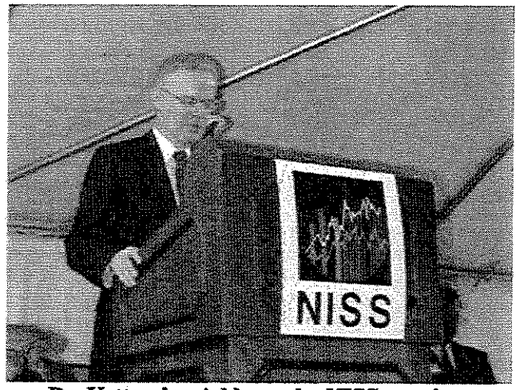
What do you think about the interactions between the academia statisticians and the industrial statisticians? There seems to be a gap in between. Of course, the old Bell Labs was kind of a special case.

Generally speaking, I think the gap you speak of has widened. This is because industry has become so short-term focused that fewer and

deeper questions that appeal to academic statisticians (not to mention industrial ones!). Researchers in academia have their own pressures. The drive for tenure may discourage junior professors from getting involved in the messy problems that they will find in industry. Notwithstanding these realities I still believe that industrial-academic collaborations can be enormously satisfying and productive, and I hope we don't give up on them.

In your opinion, how should association best serve their members? On the other hand, how should members best contribute to their associations?

The ASA has a heavy responsibility to stay in close touch with its diverse membership and to be responsive to its needs. Actually, I think it does a very good job of this overall: the meetings are well done, the publications are of the highest quality, and there is a growing array of electronic and other services that we would not have without the efforts of ASA. Sometimes I think we forget that ASA is us—it belongs to the members and the members control its effectiveness through their personal efforts and the people they elect to its offices.



Dr. Kettenring Addressed a NISS meeting.

ASA is just one of several important parts of the infrastructure that supports us as professional statisticians. The ICSA is part of it, too, as well as the IMS, ENAR, WNAR, and many other groups. I feel very strongly that all of us have a responsibility to contribute to this infrastructure, not just take advantage of it. This might happen by working on a committee of a local chapter or a section of ASA, refereeing papers, becoming a program officer at NSF, helping to organize a workshop at NISS, or serving as an officer of the ICSA. There are lots of ways. They help the profession and they help us as individuals to grow professionally.

Statisticians from Bell Labs had the opportunities to practice intensive data analyses and, therefore, learned lots of data analytical skills. However, that is not the norm among the practitioners in drug or financial business industries. The statisticians in academia seem to focus more on theory than practice. What are your thoughts about this? Is this an important issue for concern?

I had the privilege and luxury of working at Bell Labs during a very special time. Not only was I surrounded by great statisticians — Gnanadesikan, Mallows, Tukey, Wilk, and many others — but also we were given rather free rein to advance statistical methodology and to bore deeply into fascinating data analysis problems. Serious data analysis was appreciated, even admired, all the way up the management chain. I don't think it is possible to replicate that environment in today's world. However, I see no good reason why data analysis cannot be a treasured skill and a competitive advantage for companies today and especially in the industries that you singled out in your question.

Whenever I wanted to recruit statisticians, I would receive many resumes of non-Caucasian statisticians. I remember someone raised the question previously, "Where are the American statisticians?" Should the statistical Associations address this issue? If yes, how?

I know that NSF is very concerned about the pipeline for engineers and scientists in the U.S.

I know that NSF is very concerned about the pipeline for engineers and scientists in the U.S. Too many talented young people find these fields to be unattractive. Statistics is part of this picture. We need to figure out better ways of conveying the importance and excitement of statistics to students. The ASA, and other parts of our infrastructure, can provide leadership and organization of these efforts. At the same time, we shouldn't forget the many nice things that have already happened. For example, the ASA has been a major contributor to the development of new instructional materials for middle and high school students. By the way, this problem of attracting students into statistics is a natural one for us to work on in collaboration with other organizations, including the AAAS.

In my years of working experience, I see statisticians almost always serve as the supporting role. What do you think about that? Or how should statisticians do to take a more leading role?

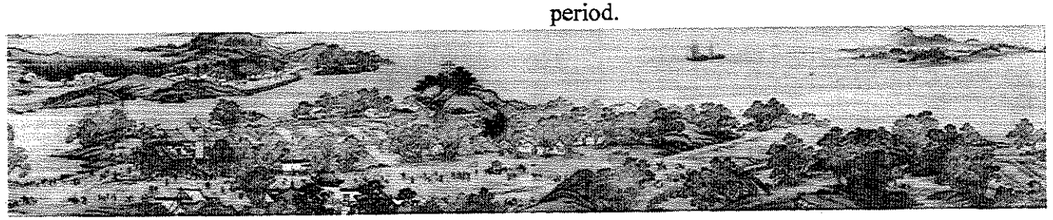
I don't have too much sympathy for the situation you describe. Obviously, we do have an important supportive and collaborative role to play. That does not mean subservient or second class—or at least it shouldn't. In the past we have had great role models such as Professors John Tukey and Fred Mosteller. Take a look at the many national problems that they had a key hand in. I am sure their involvement was as much for their leadership and clarity of thinking as it was for the specific statistical advice they had to offer. While few of us are as talented as Tukey and Mosteller, we can all "earn our stripes" as effective team members and, with that, respect and increased responsibility should follow in due course.

Applied statistics is becoming more and more computational and descriptive. Those beautiful theories we learned at school hardly get used. Should we take that into consideration in teaching of statistics? Or what direction should statistics go in your opinion?

With the increasing complexity of problems and data, one can hardly do statistical analysis without the help of computers. Data and computers are more tightly coupled than ever for statisticians. This should be reflected in our courses and curricula. Whenever feasible, one should use real data in the classroom and expose students to real problems. I recently visited a graduate school that offers data analysis only after all its methods and theory courses. I'd suggest either reversing the order or integrating data analysis into the other courses. This would help both motivation and understanding. With modern computing technologies, this can be done much easier than before.

Finally, how would you advice the younger generation of statisticians in term of steering their future directions?

I think the first job is really important. The ideal situation, in my view, is one that offers a rich set of technical challenges and talented colleagues in an environment that encourages quality work, collaboration, and professional growth. New Ph.D.s may want to consider post-doctoral appointments. For example, NISS and SAMSI offer some terrific opportunities of this kind. One caution: don't choose a position based solely on salary. Not only are there many other dimensions to think about but also, even from a strictly financial perspective, it is smarter to size up what the rewards can be over an extended period.



Controversial Statistical Issues

Bioinformatics (part I): Genomic Expression Data

Special Topic Editor: Sue-Jane Wang, Ph.D.

*Considerations of Genomics Drug Trials: Study Objectives, Replicates, Replication, Prospective Design and Its Prospect**

Sue-Jane Wang, Ph.D.

Introduction

The use of genomics/genetics to identify the pharmacologic effect has recently been widely explored. Many authors have tried to define pharmacogenomics or pharmacogenetics. Among the various definitions, here I list one defined by The European Agency for the Evaluation of Medicinal Products (EMEA): Pharmacogenomics applies genomic information to drug design, discovery and clinical development, reflecting the state or responses at cellular, tissue, individual or population levels. In this commentary, Genomics Drug Trial (GDT) will be referred to as "use of genomic technology to identify molecular signals including transcription or proteomic profiling in complex biological mixtures for use as genomic markers of disease, as pharmacogenomic markers of drug exposure or of drug response including efficacy and toxicity."

Traditionally, the pharmacologic effect of an experimental therapy (treatment) investigated from a randomized, double-blind, well-controlled clinical trial serves to weigh the observed clinical improvement or less deleterious effect in the targeted patient population described by its primary efficacy outcome against the drug related toxicity. Because of the desire to better assess new pharmacologic effect or to better quantify its effect in the more relevant patient population, the genomic efficacy endpoint, in addition to conventional primary clinical efficacy endpoint, has gradually received attention in the drug development program. As the scientific community is still trying to characterize the knowledge extracted from gene expression profiles and the findings may not be clearly replicated, use of individual patient expression profiles to interpret the pharmacologic effect is still in its infancy. The study objectives in a GDT is therefore viewed as exploratory, although they aim to embody the hypothesis test analysis as a stringent means for identification of a small set of candidate genes and to advocate the hypothesis generation analysis as a screening tool for searching an important subset of genes from the examination of genome-wide scanning.

Current Study Objective(s)

At the present time, the study objectives in GDTs are generally four-fold: diagnostic biomarker for disease severity, prognostic biomarker for clinical outcome, longitudinal biomarker for clinical outcome, and possibly surrogate biomarker for drug exposure. Instead of one gene at a time or a small number of genes simultaneously studied, a concept used in the conventional genetics, a subset of genes identified from the genome-wide scan expression profiles that tell apart the normal and the diseased individuals, or the severity levels of the targeted-disease patient population served as diagnostic expression profiles of disease. When the drug response manifests from clinical outcome in the treated patients, the association of these treatment phenotypes and the expression profiles established the prognostic biomarkers work to correlate with clinical outcome. For instance, the prognostic expression profiles may contribute to whether or not a treated patient would respond to the pharmacologic intervention. For a sufficiently large GDT, it may be possible to explore the predictive biomarker that not only is prognostic in the treated subjects, but also distinguishes between

the no difference in the controlled (often placebo vehicle) group versus the difference in the treated group. The prognostic/predictive ability of gene expression profiling is thought to provide useful tools for better therapeutic progress and to seek more homogeneous patient populations, which could increase statistical power for detection of a pharmacologic effect.

A less well-documented objective is to correlate worse (or better) clinical outcome with significant changes of expression profiles over time as compared to mild or moderate clinical worsening that showed no significant changes of expression profiles over time. The exploration of genomic biomarker in pharmacologic effect time course study could provide plausible explanation of the longitudinal expression changes that impact drug response. Patients' drug exposure characterized by the AUC of plasma concentration within some time interval, a pharmacokinetic parameter, may correlate with expression change from baseline. The observed association may signify potential biomarkers for drug exposure.

The Prospect of a GDT – Replicates in and Replication of Prospective Design

The problem of human variability in disease was well presented by Sir William Osler (1849-1919). He stated that "Variability is the law of life, and as no two faces are the same, so no two bodies are alike, and no two individuals react alike, and behave alike under the abnormal conditions which we know as disease." The apparent variability in clinical symptoms among patients could generate controversy over the prescribed pharmacologic effect often presented by the averaged targeted population. Such a paradox in modern drug development comes from "Clinical trials provide evidence of efficacy and safety at *usual doses* in *populations*" on the one hand and "Physicians treat *individual* patients who can *vary widely in their response to drug therapy*" on the other.

It has been argued that the impact of pharmacogenetics on clinical practice is relatively modest and replication of the finding is scarce. Literature suggests that many initially reported associations are in fact chance findings or false positive in the statistical sense. In view of the similar role of expression signatures in drug development, the investigative finding of diagnostic, prognostic, predictive, and surrogate biomarkers might provide some basis for confirmation in a prospectively designed and conducted genomic drug clinical trial. To prospectively design a GDT, sample size estimation involves more parameters than those we have in the usual well-controlled clinical trials, as characterization of gene expression profiles needs to be accounted for through parameters such as sensitivity, specificity, positive versus negative predictive values and possibly false discovery rate of the identified genomic profiles.

Through hypothesis generation and hypothesis test analyses performed in the various development phases of a drug development program, it is generally believed that genomic markers that could be linked to disease pathophysiology may provide sound biological justification and validation, and be easier to incorporate into a GDT. A well-known example is Herceptin, which is indicated for treatment of metastatic breast cancer patients who expressed the genomic marker HER-2 protein as either a naïve or second line therapeutic agent. However, it is not as obvious when the biomarkers are identified during clinical development. The validation of these observational biomarkers without proper biology or pathway network support prompts the need for prospective confirmation possibly through an adaptive GDT design. When the false positive of the seeming genomic association is highly suspected, either caused by design limitation or attributed to observational chance finding, replication of the investigative genomic association would make it convincing by performing a large prospective GDT to support the observed pharmacogenomics association.

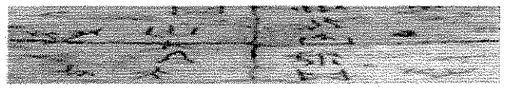
In a recent draft guidance for Industry on "Pharmacogenomic Data Submissions" published in November 2003, the U.S. Food and Drug Administration requires formal review of investigational new drugs or new drug applications should the pharmacogenomic findings be used for decision making in clinical trial or in a preclinical safety study, or used by drug sponsor to support arguments on safety, efficacy, dosing or pharmacology, or intended to be used in the drug

label or to support drug approval. Voluntary genomic data submission is highly encouraged. The encouragement lays grounds to more research work on the genomic biomarkers that are not known valid or not well-established in the literature. The voluntary genomic study is usually a substudy of a phase II double-blind, randomized, well-controlled clinical trial. Thus, the patient populations in such a study are voluntary, small or tend to be limited. Although such a study can be prospectively designed, the findings are often considered observational under the voluntary submission. One of the rationales of the Agency's initiative on the voluntary submission is to support critical path research evaluation. It is hoped that incorporation of genomics/genetic information using high throughput technology may enhance the phase II/III clinical drug development and help seek the responder patient population.

The joint learning experience of the U.S. Food and Drug Administration among the pharmaceutical industry, the academia, and the government agencies through evaluation of the voluntary genomic data submission may bring the innovative development of new treatments by moving from the current empirical process to the mechanism-based process that is hypothesis driven. It is also anticipated that if the pharmacogenomic outlook is promising, a lower cost, faster process resulting in more effective and less toxic drugs for a smaller patient population is likely to be a reality.

Acknowledgement: PhRMA/FDA Genomics (Microarray) Biostatistics Workshop, April 15-16, 2004, University System of Maryland, Rockville, MD. Statistical Analysis of Clinical Genomic and Proteomic Data, 40th DIA Annual Meeting, June 16, 2004, Washington DC.

Sue-Jane Wang, Lead Senior Mathematical Statistical, FDA Inter-Center Pharmacogenomics/ Pharmacogenetics Initiative Working Group
*The views presented in this article do not represent those of the U.S. Food and Drug Administration. Email: WANGS@cder.fda.gov



To Combine or Not to Combine

Cheng-Yan Kao, D. Frank Hsu, Han-Yu Chuang, Chi-Ying F. Huang, Kuang-Chi Chen

Life sciences are currently at the center of the information revolution. The development of techniques and tools in genomes produce the collection of biological information at an unprecedented level of detail and in extremely large quantities. The human genome project has shown us the dramatic changes in the use of information, increased computing power and new algorithms to analyze whole genomes. In this information revolution, data gathering capabilities have greatly surpassed data analysis techniques. Researchers are devoted to discovering more effective and appropriate approaches in computational genomics. Bioinformatics is an emerging discipline situated at the interface between computing information science and biological sciences such as molecular biology and genetics (Draghici (2003)). Essentially, bioinformatics is the science of refining biological information into biological knowledge using computing methods and information technology. The challenge of bioinformatics is to analyze, interpret and understand all data that are being produced (Eisenberg et al. (2000); Lockhart and Winzeler (2000)). To biological scientists, the challenge is to use the large-scale data to discover and understand fundamental biological phenomena. At the same time, the challenge faced by computer scientists is to develop new systems and techniques to support such discoveries.

Typically, hundreds or thousands of gene expression data are produced simultaneously from a small number of samples in cDNA microarray or oligonucleotide array experiments (Schena (1999)). In large-scale gene expression data, researchers are challenged by a variety of issues: noise, normalization, experimental design, large number of genes, significance, array quality assessment, biological factors, etc (Draghici (2003)). For a specific experiment, only some of the genes will be useful to differentiate samples among different classes, while many other genes are irrelevant to

this task. Those irrelevant genes not only introduce unnecessary noise to data analysis, but also increase the computational difficulties in other tasks such as clustering. The major interesting issue in bioinformatics is to select a list of differentially informative genes from array data. Several methods of identifying significant genes have been proposed in the past. However, due to the multi-faceted nature of biological activity or data structure, it is often pointed out that no single method is perfect in every study. The outcomes of different methods may differ substantially and the discordance causes difficulties in the interpretation of the dataset. Moreover, it is often uncertain which method should be applied to new unknown datasets. Thus, combining meaningful results from different methods seems to be a promising approach.

Combination methods and data fusion have been studied in a variety of different application domains such as information retrieval (IR) (Hsu, Shapiro and Taksa (2002); Ibraev, Ng and Kantor (2001)), pattern recognition (Xu, Krzyzak and Suen (1992)), and molecular similarity searching (Ginn, Willett, and Bradshaw (2000)). In Chuang et al. (2004), we apply method combination and data fusion to analyze large dataset from array experiments. Experiments using average combination and precision evaluation (Hsu, Shapiro and Taksa (2002)) indicate that no single gene selection method performs effectively across different datasets. In most our experiments, genes selected by combination methods are as informative as each of individual methods. As the number of methods combined increases, the minimum precision with the same number of methods combined increases, but the maximum drops when the number used is greater than two. Moreover, the combination of many of the heterogeneous and well-performing methods may achieve the best performance. Furthermore, by investigating the rank-score graph defined by Hsu, Shapiro and Taksa (2002) of original gene selection methods, the longer distance between two methods represents the more different between these two methods. All of these evidences indicate that method combination will generally result in a level of performance that is at least as good as the best individual method which often varies from one target data to another in an unpredictable manner. The use of fusion rules to combine the results of multiple methods provides not only a simple and effective approach to the array data analysis, but also a robust and more consistent level of searching performance than a single method, although there is yet to be found any obvious predictive mechanism for identifying an optimal combination.

Combination of multiple methods have been shown to be effective (Chuang, et al. (2004); Ginn, Willett and Bradshaw (2002); Hsu, Shapiro and Taksa (2002); Hsu and Palumbo (2004); Ibraev, Ng and Kantor (2001)). However, the question of why and how combination should be done still remains largely unanswered. The rank-score graph defined by Hsu, Shapiro and Taksa (2002) provides a better understanding of the data fusion phenomena. In addition, a rank-based combination performs better than a score-based combination, in many cases, especially for methods with heterogeneous and incompatible rank-score graphs. Firstly, as researchers are more likely to be concerned with number of most informative genes to the target dataset, it seems logical to consider the rank positions of genes irrespective of their scores. Secondly, despite having the normalized scores, the distributions of normalized scores of selection methods are markedly different and may not be directly comparable, with the possibility of biasing the fusion rule (Ginn, Willett and Bradshaw (2002)). Moreover, a weighted rank combination of IR systems improves the effectiveness of combination methods (Hsu and Palumbo (2004); Vogt and Cottrell (1999), and Ibraev, Ng and Kanto (2001)). Therefore, efficient and robust combination methods are highly desired.

A proper gene selection method considering of the property and structure of data is also considerable. Despite the best individual method which often varies from one target data to another in an unpredictable manner, it is highly desirable to select proper methods, which are most likely to have better informative and significant performance individually, into the combination list. It is not helpful to combine improper or worse effective methods with another. The choices of selection methods rely heavily on the application domain they are applied to. If heterogeneous methods are complement to one another, then

combining heterogeneous and well-performing methods might achieve the best precision. Furthermore, in order to select really informative genes rather than top n genes consideration, the critical value above some threshold of score or the significance level might need to be concerned. An adjusted p-value in a multiple testing situation should be also considered.

The proposed method can be adapted to different application domains which may call for different feature selections and different combination algorithms. Future work will explore other ways to combine different feature selection methods by different fusion rules. In a closely related study, Hsu and Palumbo (2004) consider the rank space as a Cayley graph and examine how data fusion and method combination work in that graphical model. One of the long-term goals is to construct a system which can learn from the environments and phenomena in its application domain. The system can then evolve to become a more intelligent expert system in that particular application domain.

Acknowledgements

This work was supported by grants from the National Science Council (NSC92-3112-B-005) and National Health Research Institutes to C.F.H.

References

Chuang, H.Y., Liu, H., Brown, S., McNunn-Coffran, C., Kao, C.Y. and Hsu, D.F. (2004). Identifying significant genes from microarray data. *Proceedings of BIBE'04*, IEEE CS Press.

Chuang, H.Y., Liu, H., Chen, F.A., Kao, C.Y. and Hsu, D.F. (2004). Combination methods in microarray analysis. *Proceedings of I-SPAN'04*, IEEE CS Press.

Draghici, S. (2003). *Data Analysis for DNA Microarrays*. Chapman & Hall/CRC.

Eisenberg, D., Marcotte, E.M., Xenarios, I. and Yeates, T.O. (2000). Protein function in the post-genomic era. *Nature* **405**, 823-826.

Ginn, C.M.R., Willett, P. and Bradshaw, J. (2000). Combination of molecular similarity measures using data fusion. *Prespectives in Drug Discovery and Design* **20**, 1-16.

Hsu, D.F. and Palumbo, A. (2004). A study of data fusion in Cayley graph $G(S_n, P_n)$. *Proceedings of I-SPAN'04*, IEEE CS Press.

Hsu, D.F., Shapiro, J. and Taksa, I. (2002). Methods of data fusion in information retrieval: rank vs. score combination. *DIMACS Technical Report* 2002-58.

Ibraev, U., Ng, K.B. and Kantor, P.B. (2001). Counter-intuitive cases of data fusion in information retrieval. *Rutgers University Technical Report*.

Lockhart, D.J. and Winzeler, E.A. (2000). Genomics, gene expression and DNA arrays. *Nature* **405**, 827-836.

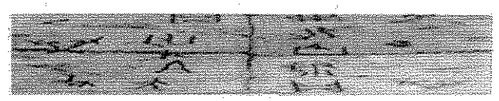
Schena, M. (1999). DNA Microarrays: A Practical Approach. *Practical Approach Series 205*. Oxford University Press, Oxford, UK.

Vogt, C.C. and Cottrell, G.W. (1999). Fusion via a linear combination of scores. *Information Retrieval* **1**, 151-173.

Xu, L., Krzyzak, A. and Suen, C.Y. (1992). Methods of combining multiple classifiers and their application to handwriting recognition. *IEEE Trans SMC* **22**, 418-435.

Cheng-Yan Kao[1], D. Frank Hsu[2*], Han-Yu Chuang[1], Chi-Ying F. Huang[1,3] and Kuang-Chi Chen[3*]

[1] National Taiwan University,
[2] Fordham University, New York
[3] National Health Research Institutes, Taipei
* Correspondent authors: chichen@nhri.org.tw, hsu@cis.fordham.edu



Statistical Aspects of Class Prediction for DNA Microarray Studies

Richard Simon, Ph.D.

Introduction

DNA microarrays are assays that provide estimates of the level of expression for thousands of genes in a specimen. The assay is based on the messenger RNA extracted from a collection of cells. The set genes for which the assay provides information may consist of all the genes of the organism. For a non-technical introduction to the biology of gene expression and the technology of DNA microarrays, see Simon et al. (2003a).

The development of DNA microarrays has created important new challenges for biomedical scientists and for statisticians. The challenges for biomedical scientists relate to learning enough about the design and analysis of DNA microarray based studies to utilize this breakthrough technology. To some extent, biomedical scientists meet this

challenge by initiating collaborations, but effective collaboration requires the scientist to understand enough about the domain of his/her collaborator to interact effectively. The challenges for statisticians involve addressing new statistical issues and learning enough about the biomedical application in order to interact as scientific collaborators and not as technical service providers.

Because microarrays have led to so much activity in several disciplines, they have also brought about many myths and misconceptions. Here, I will try to debunk some of these myths and highlight some sound statistical approaches for dealing with multiplicity issues that arise in the context of diagnostic or prognostic prediction using DNA microarray data. The focus will be on principles here, because space does not allow a comprehensive review of the numerous methods of design and analysis available.

DNA Microarray Studies Need Clear Objectives

It is a myth that DNA microarray studies involve looking for interesting patterns in data archives that provide answers to un-asked questions. Most successful microarray studies have clear objectives, although the objectives do not represent gene specific hypotheses. The objectives determine the specimens to be collected, the experimental interventions to perform and the analysis plan. There are very few examples of successful microarray studies based on data archives.

One type of objective commonly encountered in DNA microarray experiments is *class comparison*; that is, identification of genes differentially expressed among pre-defined classes of samples. A related, but distinct, objective that is often relevant for medical studies is *class prediction*; that is development of a function that can accurately predict the biologic group, diagnostic category or prognostic stage of a patient based on an expression profile of the diseased tissue from that patient.

Class discovery is fundamentally different than class comparison or class prediction in that no classes are pre-defined. Cluster analysis is useful for class discovery problems, but not usually for class comparison or class prediction. Supervised methods that utilize information about which samples are from which class are more effective for class comparison and class prediction studies. Here, we will focus on class prediction. Statistical methods for class comparison and class discovery are reviewed in (Simon et al. 2003a).

Prediction is not Inference

Most statistical methods were developed for inference, and are not really relevant for prediction. For example, we are not interested in which covariates are significant or which are significant conditional on which other covariates are in the model. Statistical significance is not the issue, nor is estimation of regression coefficients; the issue is prediction. Similarly, one should not get distracted by questions of model goodness of fit nor of uniqueness of the covariates included in the model. When the number of candidate predictors (p) is orders of magnitude greater than the number of cases (n), it is not feasible to ensure that a model is correct nor unique. It is also not feasible to assure that one has the most parsimonious model. It may be possible, however, to identify a model that predicts accurately for independent data, and that should be the focus.

Feature Selection for Class Prediction

It is well known from the theory of linear regression that including noise variables in the predictive model increases the mean-square error of prediction. A noise variable is a variable that is not related to the thing being predicted. Feature selection is generally important in microarray studies because the number of noise genes may be orders of magnitude greater than the number of informative genes. The influence of the genes that actually distinguish the classes may be lost if too many noise variables are included in the model.

A commonly used method feature selection is to identify the genes that are differentially expressed among the classes when considered individually. For example, if there are two classes, one can compute a t-test or a Mann-Whitney test for each gene. Selecting the genes that are significantly differentially expressed at a specified nominal significance level makes more sense than selecting a fixed number of genes. The stringency of the

significance level used controls the number of genes that are included in the model. Issues of multiple testing or false positives are not really relevant, however, because the objective is just to select features for inclusion in the model; no particular claim is made about the selected genes. Similarly, it doesn't really matter whether the assumptions of the t-test are strictly satisfied, because the p values are merely used as a convenient index for selecting genes.

Several authors have developed methods to identify optimal sets of genes which together provide good discrimination of the classes. Some methods are very computationally intensive and use genetic algorithms to try to identify optimal gene sets for classification. Most of these methods are presented without a comparison to simpler methods of feature selection and for some it is not clear whether the classification error rates are properly cross-validated. Consequently, it is not clear whether the increased computational effort of these methods is warranted.

Some studies use linear combinations of gene expression values as predictors. Using principal components provides a vast reduction in the dimension of the expression data, but has two serious limitations. One is that the weights given to each gene in the principal components are based on variation among samples without regard to classes. Hence, the principal components are not necessarily good predictors. The second problem is that the principal components have weights for all of the genes. Consequently, prediction using principal components requires that expression of all the genes be measured.

Selecting a Class Prediction Model

The second main component of a class predictor is complete specification of the mathematical function that will provide a predicted class label for any given expression vector x. There are many kinds of predictor functions. A linear discriminant is a function

$$l(\underline{x}) = \sum_{i \in F} w_i x_i \qquad (1)$$

where x_i denotes the expression measurement for the i'th gene, w_i is the weight given to that gene, and the summation is over the set F of features (genes) selected for inclusion in the class predictor. For a two-class problem, there is a threshold value d, and a sample with expression profile defined by a vector $\underline{x}$ of values is predicted to be in class 1 or class 2 depending on whether $l(\underline{x})$ as computed from equation (1) is less than the threshold d or greater than d respectively.

There are a large number of class predictors based on linear discriminants of the form shown in (1). They differ with regard to how the weights are determined. The oldest form of linear discriminant is Fisher's linear discriminant (Fisher 1936). Fisher LDA can be viewed as Bayesian classification with equal priors and multivariate normal distributions with a common covariance matrix (Simon et al. 2003). The study by Dudoit et al. (2002) indicated that Fisher LDA did not perform well unless the number of selected genes was small relative to the number of samples. The reason is that in general there are too many correlations to estimate and the method tends to be un-stable and over-fit the data.

Diagonal linear discriminant analysis (DLDA) is a special case of Fisher LDA in which it is assumed that there is no correlation among genes. By ignoring the correlations, one avoids having to estimate many parameters, and obtains a method which performs better when the number of samples is small. DLDA is essentially the same as what has been called the naïve Bayes classifier. Golub's weighted voting method (1999) and the Compound Covariate Predictor of Radmacher et al. (2002) are similar to diagonal linear discriminant analysis and tend to perform very well when the number of samples is small. They compute the weights based on the univariate prediction strength of individual genes and ignore correlations among the genes. They can be viewed as weighted voting among univariate classifiers.

Inner-product kernel support vector machines do classification using a linear predictor of the form of equation (1). There are, in fact, an infinite number of w vectors which completely separate the two classes in a set of training data when p>n. A support vector machine with inner-product kernel selects the w vector which separates the classes with a maximum margin between the separating

hyperplane and samples in the classes. The samples that are closest to the separating hyperplane are called support vectors. The support vector machine can be expressed as finding the hyperplane that separates the expression vectors with a specified margin and minimizes the length of the weight vector. Hence, it provides a regularized solution to the problem, in spirit similar to ridge regression for regression problems. Although there are more complex forms of support vector machines, since p>>n problems are always linearly separable, there is little reason to use other kernels. The motivation for selecting a kernel is to project the vectors in a higher dimensional space where they will be separable.

Perceptrons are neural networks with no hidden layer and hence are linear classifiers of the form of equation (1). In the study of Dudoit et al. (2002), the simplest methods, diagonal linear discriminant analysis, and nearest neighbor classification, performed as well or better than the more complex methods. Dudoit et al. (2002) also studied some more complex methods such a Classification Trees and aggregated Classification Trees. These methods did not appear to perform any better in general than the simpler methods. Ben-Dor et al. (2000) also compared several methods on several public datasets and found that nearest neighbor classification generally performed as well or better than more complex methods.

Estimating Prediction Accuracy

Evaluation of a predictor should be based on predictive accuracy. This may seem obvious, but the mistake is often made of evaluating a predictive model based on how reproducible the features or weights are if the model were developed with independent data.

The most adequate method of estimating accuracy of a predictor is to completely specify the predictor on one set of data and to test the predictor on an independent set of data collected from different centers and assayed at different times from the training data. If all of the available data is collected from a single center and assayed at a single time, then there are important sources of error for future predictions which may not be reflected in the estimate of prediction accuracy resulting from the study.

In many studies, there will not be a fully independent set of data available for estimating prediction accuracy. It may still be useful to estimate prediction accuracy using a split sample, cross-validation or bootstrap procedure as long as the potential limitations of the estimate are recognized. The re-substitution estimate is the estimate of error rate in which the same sample is used for model development and prediction, with no use of split-samples or cross-validation. Re-substitution estimates are extremely biased for microarray data and should never be reported (Simon et al. 2003b)..

With the split sample method, the current set of samples must be partitioned into a training set and a separate test set. If possible, the test set samples should be from different centers and assayed at different times than the training set samples. The test set emulates the set of future samples for which class labels are to be predicted. Consequently the test samples cannot be used in *any* way for the development of the prediction model. This means that *the test samples* cannot be used for estimating the parameters of the model and they *cannot be used for selecting the gene set to be used in the model.* Rosenwald et al. (2002) used the split sample approach successfully in their international study of prognostic prediction for large cell lymphoma. They used two thirds of their samples as a training set. Multiple kinds of predictors were studied on the training set. When the collaborators of that study agreed on a *single fully specified prediction model*, they accessed the test set for the first time. On the test set there was no adjustment of the model or fitting of parameters. They merely used the samples in the test set to evaluate the predictions of the model that was completely specified using only the training data. With the split sample approach it is easy to estimate sensitivity and specificity for each class and to obtain confidence intervals for prediction error, sensitivity and specificity.

Cross-validation is an alternative to the split sample method of estimating prediction accuracy. With cross-validation, models are independently developed from scratch in the training sets and

used to predict for the samples in the test sets. Errors are tabulated on the test set predictions and totaled.

With leave-one-out-cross-validation (LOOCV) n different models are developed based on n training sets of (n-1) samples each. The model developed with sample i omitted is used to predict the class of sample i and the errors are totaled. The model that would be used for future predictions is one constructed using all n samples. The cross-validated error rate is an estimate of the error rate to be expected in use of this model for future samples. The n models are constructed only in order to estimate the prediction error associated with the type of model constructed.

There are several misconceptions prevalent about cross-validation. One is that one can select the features using all of the data and just use cross-validation to re-fit the model for each leave-one-out training set for the same set of genes. Ambroise and McLachlan (2002) and Simon et al. (2003b) showed that this results in an extremely biased estimate of prediction error for p>>n problems. Because p>>n, using all n samples to select the features is very different than using n-1 of the samples for most feature selection methods, and hence the greatest component of potential bias is in the feature selection.

A second misconception about cross-validation is as follows. Suppose you have a model development method containing a tuning parameter δ. You use cross-validation to estimate the prediction error for each value of δ on a grid, say $\hat{Err}(\delta_j)$ as an estimate of the true value $Err(\delta_j)$. You then select the value δ* at which $\hat{Err}(\delta_j)$ is minimized. Unfortunately, $\hat{Err}(\delta^*)$ is a biased estimate of $Err(\delta^*)$. If you want to use the cross-validation results to select your model for future use, then you must recognize that the final selection is part of your model development algorithm and it should be incorporated within each cross-validation training set. This also applies to k-fold cross-validation and to use of the bootstrap.

There are also invalid criticisms of cross-validation based on assertions that it is always possible to find a model which predicts perfectly for a training set and test set (Somorjai et al. 2003). There are actually an infinite number of linear models that predict perfectly for any training set in which p>>n but most of them will not predict perfectly for the test set. Cross validation requires that you have an algorithm which selects a single fully specified model based only on training set data. Proper use of cross validation will provide an almost unbiased estimate of the prediction error for that model development process.

Radmacher et al. (2002) have defined a paradigm for the internal validation of a class predictor. It includes generating the permutation distribution of the cross-validated mis-classification rate as a means of evaluating the prediction model.

Conclusion

The development of biotechnology has created a great need for statisticians to collaborate in the design and analysis of genomic data. There are new statistical challenges and new opportunities for statisticians to make important contributions to the understanding of disease and the development of effective therapeutics. Statisticians should retain their commitment to sound statistical principles for design and analysis but must be open to learning in depth new subject matter areas so that they can participate as full collaborators.

The BRB-ArrayTools software which I have developed in collaboration with Amy Peng Lam (Simon and Lam 2004) contains most of the analytic methods described here including proper LOOCV, k-fold cross-validation and .632 bootstrap based prediction error estimation for a wide variety of prediction models including DLDA, compound covariate predictor, nearest neighbor methods, support vector machines, shrunken centroids and random forests. It also contains tools for class comparison and class discovery as well as some of the bioconductor tools such as RMA for pre-processing of Affymetrix .CEL files. The software is available at http://linus.nci.nih.gov/brb.

References

Ambroise, C. & McLachlan, G.J. (2002). Selection bias in gene extraction on the basis of microarray gene-expression data. *Proceedings of the National Academy of Science*, **99**, 6562-66.

Ben-Dor, A., Bruhn, L., Friedman, N., Nachman, I., Schummer, M. & Yakhini, Z. (2000). Tissue classification with gene expression profiles. *Journal of Computational Biololgy*, **7**, 559-584.

Dudoit, S., Fridlyand, J. & Speed, T.P. (2002). Comparison of discrimination methods for the classification of tumors using gene expression data. *Journal of the American Statistical Association*, **97**, 77-87.

Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., Bloomfield, C.D. & Lander, E.S. (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*, **286**, 531-537.

Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U., Speed, T.P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4:249-264.

Radmacher, M.D., McShane, L.M. & Simon, R. (2002). A paradigm for class prediction using gene expression profiles. *Journal of Computational Biolology*, **9**, 505-12.

Rosenwald, A., Wright, G., Chan, W.C., Connors, J.M., Campo, E., Fisher, R.I. & al., e. (2002). The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *New England Journal of Medicine*, **346**, 1937-47.

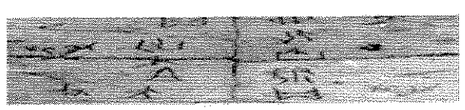
Simon R, Korn E, McShane L, Radmacher M, Wright G, Zhao Y. (2003a). *Design and Analysis of DNA Microarray Investigations*, Springer-Verlag New York.

Simon, R., Radmacher, M.D., Dobbin, K. & McShane, L.M. (2003b). Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. *Journal of the National Cancer Institute*, **95**, 14-18.

Simon, R. & Peng-Lam, A. (2004). BRB-ArrayTools Users Guide, Version 3.2. *Technical Report, Biometric Research Branch, National Cancer Institute, http://linus.nci.nih.gov/brb*.

Somorjai RL, Dolenko B, Baumgartner R. (2003). Class prediction and discovery using gene microarray and proteomics mass spectroscopy data: curses, caveats, cautions. Bioinformatics 19:1484-1491.

Richard Simon
National Cancer Institute
9000 Rockville Pike,
MSC 7434, Bethesda MD 20892
rsimon@nih.gov



Comments on Validation of Normalization and Background Correction for Microarray Data

Robert R. Delongchamp, Ph.D.
Ralph Kodell, Ph.D.

Introduction

The cDNA microarray has the potential to be an important tool in drug discovery, and in the evaluation of safety and efficacy. Before array technology can reach its full potential for providing reliable genomic information to bioinformatics efforts in the regulatory decision-making process, a number of problematic issues will have to be resolved. At the FDA-PhRMA workshop held April 15-16, 2004 in Rockville, MD experts from the FDA, NIH, the pharmaceutical industry and academia gathered to discuss in the context of case studies the pressing biostatistical challenges presented by microarrays. A common need identified by the pre-clinical and clinical breakout groups was the validation of platforms and procedures so that the information collected by different investigators using different platforms can be exchanged.

The steps involved in processing a tissue sample in a microarray study include RNA extraction, transcription into a labeled DNA, and hybridization to the arrayed DNA (Nguyen, *et al.* 2002). Systematic effects are introduced in this process. Fundamental to any statistical analysis of microarray data is the 'normalization' of the array data to eliminate these biases. Usually, an adjustment for background precedes normalization or is implicit in the normalization procedure. There is no gold standard for data normalization, although regression methods such as LOESS have been popularized for dealing with specific sources of bias, such as spatial heterogeneity, differential background and signal saturation (Quackenbush 2002, Yang, *et al.* 2002).

As was pointed out at the FDA-PhRMA workshop, the normalization step can have a profound impact on the set of genes ultimately classified as differentially expressed via formal statistical analyses. Although some degree of misclassification of individual genes with respect to over-, under-, or normal expression is inevitable, careful normalization can help to minimize such misclassification. This is clearly important in the potential use of microarrays as tools and sources of data for class prediction, e.g., cancer vs non-cancer. It is even more important if microarrays are to be useful for the study of mechanisms of action. This article will discuss some of the problematic issues accompanying data normalization, including background adjustment.

Problems with normalization and background subtraction

The observed data from a study are label intensities for a set of *m* probes (genes), measured on *n* arrays. The intensities indirectly measure mRNA concentrations in the tissue samples (gene expression). In addition to differences in mRNA concentrations across samples (treatment effects and biological variation), intensities differ across arrays because of phenomena such as cross-hybridization, background intensity, arbitrary amplification of measurements, saturation, etc. (technical variation) (Churchill 2002, Yang and Speed 2002). The technical sources of variation are large enough to overwhelm most treatment effects, especially with typical sample sizes. Some of the technical sources induce systematic changes in intensities that can be observed within specified subsets of probes on an array. Such systematic changes can be estimated and subsequently eliminated. In the jargon, the observed intensities are 'normalized'. Some form of normalization is invariably applied either prior to analyses for treatment effects or as part of the analyses (Parrish and Delongchamp 2004). While the desirability of normalization is universally recognized, there is no consensus as to a method. Because a lion's share of variation is usually removed by normalization, the inferred treatment effects will depend upon the normalization method.

A need for normalization originates with the underlying sampling procedure. An assay for the gene expression in a tissue sample begins with a fixed aliquot from the total mRNA that was extracted from the tissue. As a consequence, only the proportions of the total mRNA can be ascertained for the constituent mRNA species in the aliquot. In effect, one cannot distinguish whether a given mRNA increased in concentration or whether the other mRNA species decreased. This identifiability problem can not be resolved without additional information, which is usually provided as an assertion. For examples, all samples have the same concentration of total mRNA, or all samples have the same concentration of specified mRNAs (genes). Such constraints are implemented as global adjustments, i.e., all intensities, which are recorded for the aliquot, receive the same adjustment.

There are several additional technical sources of variation, which are superimposed upon the identifiability problem. They differ conceptually because these effects can be estimated through replication and because these effects do not necessarily apply to every probe on an array. In our work, we favor experimental designs that directly estimate these effects. That is, each experimental unit is measured on two arrays such that probe-specific batch effects, dye effects, and array effects can be partitioned in an analysis of variance (Chen, *et al.* 2004, Desai, *et al.* 2004). Since this approach uses two arrays per experimental unit, it is somewhat hard to sell. The identifiability problem remains and so the need for a global normalization. However, such designs are expected to eliminate the other systematic technical effects. They often work as intended, but not always. In which case, the value of the more elaborate design is questionable.

Other normalization approaches, which assume relationships to adjust intensities within arrays, also can eliminate systematic effects. They rely entirely on the validity of the assertions that achieve identifiability. In as much as the assertions are wrong, then the normalized intensities will be biased. In general, there is a trade-off between variance and bias. Normalization methods that can remove most of the variation incur a potential for large biases. The actual bias also depends upon the number of affected genes and the magnitudes of the treatment effects (Parrish and Delongchamp 2004). This becomes especially problematic when

treatments yield substantial changes in the expression profile.

The observed intensity includes background. The background contribution should be small, and it is of little consequence for genes that are highly expressed. However, background makes it difficult to distinguish low levels of expression from no expression tempting one to subtract an estimate of background from the intensity. The available estimates are problematic. Intensities adjacent to the spot are sometimes used to estimate background. These estimates have two problems. First, they cannot measure cross-hybridization to the probe from other cDNA species, which is an important source of background. That is, background measured from an adjacent surface, which is not spotted with cDNA, is an irrelevant estimate of cross-hybridization. Second, there can be 'over shine'. That is, a measurement in the area adjacent to the spot can pick up signal scattered from the spot. An alternative is to specifically spot foreign DNA probes to measure cross-hybridization. This estimate does not necessarily represent the average cross-hybridization very well. Further, these estimates are usually based upon a few spots and they are subject to variation. It is not clear that any bias removed by such a background subtraction compensates for the increased variation. Furthermore, background intensities can exceed the observed intensity for a gene so that a background subtraction tends to preclude interesting data transformations, e.g., logarithms. Most likely, genes yielding negative adjusted intensities are not expressed or minimally expressed, and it is a common practice to delete these genes from further consideration. However, such deletions preclude a straightforward comparison with other arrays where these genes might be 'turned on' by a treatment.

Some normalization methods adjust for apparent trends in plots of differences versus sums of log-transformed intensities from two arrays. This can mitigate against artifacts that arise when arrays have different background contributions. Regardless of how one deals with background, it makes the interpretation of observed expression differences at low expression levels problematic.

In summary, intensities recorded in microarray studies require normalization. They can be normalized either before statistical analyses or as a component of the analyses. The intensities also contain background and at least for genes having low expression levels, background effects can be consequential. The inferred treatment effects are dependent on the normalization method and background correction. Hence, the scientific and regulatory communities need sound criteria for evaluating normalization procedures. We believe that this necessitates abandonment of *ad hoc* approaches to normalization and we suggest that a more valid approach is to directly estimate any effects that normalization purports to remove. This requires more elaborate experimental designs using more arrays than is typical of current practice. To be successful, this approach requires identification of important sources of technical variation and the implementation of an experimental design that adequately copes with these sources. A global normalization will still be required to convert the 'proportions' assayed in aliquots to concentrations in the tissue samples. In as much as the conversion factor relies on *a priori* assumptions, it will be a weakness of inferences based upon microarray technologies.

References

Chen, J. J., Delongchamp, R. R., Chen-An, T., Desai, V., and Fuscoe, J. (2004), Sources of variation in microarray gene expression data, *Bioinformatics*, (in press).

Churchill, G. A. (2002), Fundamentals of experimental design for cdna microarrays, *Nature Genetics.*, **32**, 490-495.

Desai, V., Moland, C. L., Branham, W. S., Duffy, P. H., Delongchamp, R. R., Fang, H., Peterson, C. A., Beggs, M. L., Moore, M. M., and Fuscoe, J. C. (2004), Microarray analysis to examine changes in expression level of genes as a function of time of day in Fischer-344 rat liver., *Mutation Research*, (in press).

Nguyen, D. V., Arpat, A. B., Wang, N., and Carroll, R. J. (2002), DNA microarray experiments: Biological and technological aspects, *Biometrics*, **58**, 701-717.

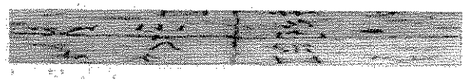
Parrish, R. S., and Delongchamp, R. R. (2004), "Normalization," in *DNA microarrays and statistical genomic techniques: Design, analysis, and interpretation of experiments*, eds. D. B. Allison, G. P. Page, T. M. Beasley and J. W. Edwards, New York: Marcel Dekker, Inc., p. (in press).

Quackenbush, J. (2002), Microarray data normalization and transformation, *Nature Genetics*, **32**, 496-501.

Yang, Y. H., Dudoit, S., Luu, P., Lin, D. M., Peng, V., Ngai, J., and Speed, T. P. (2002), Normalization for cDNA microarray data: A robust composite method addressing single and multiple slide systematic variation, *Nucleic Acids Research.*, **30**, e15:11-e15:11.

Yang, Y. H., and Speed, T. (2002), Design issues for cDNA microarray experiments, *Nature Reviews Genetics.*, **3**, 579-588.

Robert R. Delongchamp and Ralph L. Kodell
Division of Biometry and Risk Assessment
National Center for Toxicological Research
FDA, Jefferson, Arkansas 72079 USA
Email: rdelongchamp@nctr.fda.gov
rkodell@nctr.fda.gov



Statistical Science in the Genomic Era: Where are we heading?

Jae K. Lee, Ph.D.

Accelerated by the Human Genome Project, recent advances in high-throughput biotechnologies have dramatically changed the horizon of biological and biomedical sciences. Large-screening expression profiling techniques such as DNA microarrays, mass spectrometry, and protein chips offer great promise for functional genomics and proteomics research, and have the potential to transform the diagnosis and treatment of human diseases (Sander, 2000). In particular, DNA microarray and GeneChip™ gene expression approaches are becoming increasingly more important in genome-wide biomedical studies (Golub et al., 1999; Pinkel, 2000; Scherf et al., 2000).

Analysis of such genome-wide data, however, has brought extreme challenges not only in the biological sciences but also in the statistical sciences. Indeed, fundamental difficulties exist in applying traditional statistical approaches to genome-wide expression data, namely *the multiple comparisons issue* and *the small n--large p problem* (Lee, 2002). The former problem arises because classical statistical hypothesis testing, modeling, and inference strategies are designed for studying a small number of candidate targets at a time, whereas one typically investigates tens of thousands of genes' differential expression in a single microarray study. For example, when a two-sample t-test is applied for evaluating statistical significance of thousands of genes' differential expression patterns in a microarray study, the p-values obtained from this within-gene test must be adjusted to take into account the random chance of all the candidate genes in the array data.

The latter difficulty--- *the small n--large p problem* arises due to the fact that many biological and biomedical microarray studies are performed with a small number of replicated arrays, or without replication. Unlike sequence information, gene expression data are context-dependent and offer different interpretations depending on (patient) sample condition, time point, and treatment for a single subject (Stoeckert et al., 2002). In addition to the high costs of microarray experiments, certain biological or human patient specimens are often limited, thereby necessitating that microarray studies be performed with limited replication. Consequently, one must perform statistical inference on a small number of observations (n) compared to a large number of potential predictor genes (p). The latter number is simply too large to be considered in standard statistical testing and modeling, whereas the sample size (or number of replicated arrays at each condition) of a microarray study is typically small, a few tens at most and often only one or two replicates. This presents great difficulty for the application of traditional statistical approaches, which generally require a reasonably large sample size for maximal performance. As microarray (and similar high-throughput) technology becomes an important tool in biological and biomedical investigation, the lack of appropriate statistical methods for the large-screening of microarray data with limited replication will undoubtedly become a great obstacle in genomic research in medicine.

In order to circumvent these difficulties, innovative statistical concepts and approaches have been recently introduced. For example, in statistics the family-wise error rate (FWER) has been customarily used to control for the random chance of multiple candidates by evaluating the probability that, at most, one false positive is included at a cutoff level of a statistic (Dudoit et

al., 2002). However, FWER adjustment has been found to be very conservative in microarray studies, resulting in a high false-negative error rate (Tusher et al., 2001). To avoid such a pitfall, a novel concept of statistical significance, the so-called false discovery rate (FDR) and the q-value have been suggested (Benjamini and Hochberg, 1995; Benjamini and Yekutieli, 2001; Storey, 2003). Unlike FWER, FDR controls the expected proportion of false positives among all candidate targets identified significantly by a testing criterion. Based on FDR, researchers can now assess their statistical confidence among the identified targets with a much smaller false-negative error rate. The FDR evaluation has been rapidly adopted for microarray data analysis including the SAM (Significance Analysis of Microarrays) approach (Tusher et al., 2001; Dudoit et al, 2002; Jain et al., 2003a).

Inference on small sample data sets is extremely difficult using the traditional statistical framework that requires a significantly larger number of observations than the number of potential predictors (Box and Meyer, 1986). A small sample size results in under-powered inference by most available statistical approaches. As a result, many authors have recommended that a large number of replications be used in microarray experiments (e.g., Kerr and Churchill, 2001). However, as discussed above, this is not always possible in many microarray studies. When only a small number of replicates are available (e.g. duplicate or triplicate), the use of naive, within-gene estimates of variability does not provide a reliable hypothesis-testing framework. For example, a gene may have very similar differential expression values in duplicate experiments by chance alone. Furthermore, error estimates constructed solely within genes have large variability of their own, so that these error estimates result in underpowered tests and large numbers of false positives in small sample microarray studies. This can also lead to inflated signal-to-noise ratios for genes with low but similar expression values (Lee and O'Connell, 2003).

Array experiments are frequently performed under multiple conditions or time points (Eisen et al., 1999; Scherf et al., 2000; Jain et al., 2003a). Several linear modeling approaches have been introduced to analyze array data under multiple conditions (Kerr et al., 2000; Wolfinger et al., 2001). However, in the classical linear modeling paradigm, it is extremely difficult, or impossible, to parameterize and estimate the effects of thousands of genes, their interactions, and multiple error terms for different combinations of array experimental conditions and genes. A *global* ANOVA model inference that simultaneously parameterizes the effects of all the genes and their interactions in a single linear model is restricted because inverting its corresponding design matrix X (or $X\ X^T$) of these (tens of thousands) parameters becomes computationally infeasible. Even for the *within-gene* ANOVA model inference that fits each ANOVA model gene by gene, as in Kerr and Churchill (2001), a good number of replicated arrays per condition would be required for accurate estimation of all the main and interaction parameters. Furthermore, these linear modeling approaches cannot capture heterogeneous variability from various sources of error such as sample preparation, labeling, hybridization, image processing, and each gene's innate heterogeneous biological variability under different conditions. Linear least-square estimation solutions for various error terms in a split-plot design model can often result in negative estimates of errors (Milliken and Johnson, 1984).

Therefore, novel statistical approaches are in high demand for the analysis of large screening microarray data sets with limited replication. Several attempts have been made to address this rising need. For example, a number of approaches aimed at improving estimates of variability and statistical tests of differential expression have been proposed. Nadon et al. (2001) used a smoothing-spline pooled-error method by regressing standard error estimates on the mean log intensities. Durbin et al. (2002) estimate a two-parameter variance function of mean expression intensity. Lin et al. (2003) suggested the use of a data-adapted robust estimate of array error based on a smoothing spline and standardized local median absolute deviation (MAD). Kamb and Ramaswami (2001) suggested a simple regression estimation of local variances. Huang and Pan (2002) compared some of these variance estimation methods.

The local pooled error (LPE) method, a refined error-pooling method for within-gene error, has also been introduced whereby variance estimates for genes are formed by pooling variance estimates for genes with similar expression intensities from (a small number of) replicated arrays within each experimental condition (Lee and O'Connell, 2003; Jain et al., 2003a). This extension is possible under the assumption that the majority of genes in a microarray (or genome-wide) study are unregulated (not differentially expressed) between the different conditions of an array study. This assumption is generally true in most microarray studies because researchers are interested in, and experiment with, subtle expression differences of a small proportion of genes (benign vs. malignant tumor samples, control vs. disease groups, and prior vs. post treatment on the same biological stimuli). Therefore, LPE and similar error-pooling approaches avoid the small n--large p problem by leveraging the observations that many unregulated genes with similar expression intensity values often show similar array-experimental variability (Lee, 2002; Lee and O'Connell, 2003). In particular, the LPE variance function borrows strength across genes in order to improve reliability of variance estimates, from which the resulting LPE tests can dramatically improve the statistical power for identifying differential expressions with a small number of replicates. Note that this is conceptually similar to the empirical Bayes methods of Lonnstedt and Speed (2002) and Baldi and Long (2001), which shrink the within-gene variance estimate towards an estimate including more genes, and construct signal-to-noise ratios using the shrunken variance.

In order to circumvent the restrictions of the ANOVA modeling approaches, several approaches have been developed through consideration of a mixture distribution of two or more different distributional structures that depend on each gene's differential expression status, e.g. differential or equivalent expression patterns. Broet et al. (2002) proposed a multiple-component normal mixture model for the difference between two conditions in a Bayesian framework, while Allison et al. (2002) proposed a multiple component Beta mixture model for p-values, arising from testing the difference between two conditions. Newton and Kendziorski (2003) suggested that a semi-parametric hierarchical mixture model is sensitive and flexible in detecting differential expression with heterogeneous variability. Efron et al. (2001) and Efron and Tibshirani (2002) have introduced a nonparametric mixture model using an empirical Bayes approach to capture the prior specifications of differentially and non-differentially expressed genes. Lee et al. (2001) and Cho and Lee (2004) have considered a Bayesian heterogeneous error model (HEM). HEM considers two hierarchical layers, one for array experimental (technical) error variability and the other for biological error variability, from which HEM allows one to simultaneously parameterize and estimate the effects of a large number of genes, interactions, and heterogeneous error components for different combinations of genes and conditions. Compared to the ANOVA approaches, these Bayesian models can strengthen their statistical power by utilizing other genes' information implicitly through their conditional distributions, or directly by prior specifications.

However, there still exist many remaining questions in statistical applications to genomic data. For example, the current FDR (or q-value) evaluation is largely based on the adjustment methods suggested by Benjamini and Hochberg (1995) and Benjamini and Yekutieli (2001) under certain dependency assumptions or by a simple permutation sampling (Storey and Tibshirani, 2003). These FDR evaluation methods are still found to be conservative and/or do not sufficiently capture the underlying biological distribution of microarray data (Jain et al., 2003b). Various parametric and non-parametric modeling assumptions and prior specifications have been used in microarray data analysis. These distributional assumptions are often found to significantly differ from the *biologically-relevant* null distribution that can be observed in real microarray data without any biological perturbation; for example, a simple random permutation will completely break the innate correlation structures of genes and biological subjects, so that the underlying null distribution for evaluation of statistical significance is inappropriate for the discovery of biologically important genes under real biological conditions. A better understanding on the distributional characteristics of microarray and other high-

throughput data, especially with a small number of replications is urgently needed in statistical modeling and inference.

Once useful statistical methodologies are introduced for genome-wide analysis, it will be critical to develop reliable and convenient software that can be easily used by both statistical and biological investigators. This computing development is equally important to the methodological development in order to provide the newly-created analysis techniques to biomedical researchers in a user-friendly format and to other statistical researchers for further validation and refinement. Statistical scientists must also play a significant role in the computing development to rigorously test and validate such software. The Open-source Bioconductor site (www.bioconductor.org) is such an initial attempt, which has become one of the most popular software depositories in bioinformatics.

The statistical methods developed for genomic and bioinformatics data should be considered carefully for whether or not they enhance the analysis of the data in practical biological and biomedical settings. Such statistical developments will significantly improve the utility of using high throughput technologies in biological and biomedical research. For example, using these analysis methods, a much larger number of biomedical conditions and mechanisms will be simultaneously explored. Based on such improved analysis capabilities, more complex array experiments will be designed to understand complex expression mechanisms of human diseases. As the statistical and biological sciences continue to build upon the work of Mendel and Fisher, the great masters of genetics and statistics, statistical researchers must understand the complex nature of the biological problems in genomics and bioinformatics and be willing to explore and try to remove unrealistic and impractical restrictions in the current statistical environment.

Acknowledgements

This study was supported by the American Cancer Society grant RSG-02-182-01-MGO.

References

Baldi, P. and Long, A. D. (2001) A Bayesian framework for the analysis of microarray expression data: regularized t-test and statistical inferences of gene changes, Bioinformatics, 17, 509-519.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society, Series B, Methodological 57:289-300.

Benjamini Y, Yekutieli D (2001) Controlling the false discovery rate in multiple hypothesis testing under dependency, Annals of Statistics, *to appear*.

Cho, H. and Lee, J.K. (2004) Hierarchical error model for analyzing gene expression data, Bioinformatics, forthcoming.

Dudoit, S., Yang, Y. H., Callow, M. J. and Speed, T. P. (2002) Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments, Statistica Sinica, 12:1, 111-139.

Durbin, B. Hardin, J., Hawkins, D. and Rocke, D. (2002). A Variance-Stabilizing Transformation for Gene-Expression Microarray Data, Bioinformatics, 18: S1, 105-110.

Efron, B and Tibshirani, R. and Storey, J.D. and Tusher, V. (2001). Empirical Bayes Analysis of a Microarray Experiment, Journal of American Statisticl Association, 96,1151-1160.

Efron, B and Tibshirani, R. (2002). Empirical Bayes Methods and False Discovery Rates for Microarrays, Genetic Epidemiology, 23, 70-86.

Eisen MB, Spellman PT, Brown PO, Botstein D (1998). Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci. Dec 8;95(25):14863-8.

Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science, Oct15;286(5439):531-7.

Ibrahim, J.G. and Chen, M.-H. and Gray, R.J. (2002). Bayesian Models for Gene Expression with DNA Microarray Data, Journal of American Statistical Association, 97, 88-99.

Kamb, A. and Ramaswami, A. (2001) A simple method for statistical analysis of intensity differences in microarray-derived gene expression data, BMC Biotechnology, 1-8.

Jain, N., Ley, K., Thatte, J., O'Connell, M., and Lee, J.K. (2003a) Local pooled error test for identifying differentially expressed genes with a small number of replicated microarrays, *Bioinformatics,* 19(15), 1945-51.

Jain, N., Ley, Haaland, P., O'Connell, M., and Lee, J.K. (2003b) Local pooled error test for a small number of replicated microarrays, Proceedings of CHI Microarray Data analysis, Sep. 22-23, Baltimore.

Kerr and Churchill (2001). Bootstrapping cluster analysis: Assessing the reliability of conclusions from microarray experiments, PNAS, 98:16, 8961-5.

Milliken GA and Johnson DE (1984). Analysis of Messy Data (vol. I). Van Nostrand Reinhold: New York

Lee, J.K., Scherf, U., Smith, LH, Tanabe, L., and Weinstein, J.N. (2001). Analysis of gene expression data of the NCI 60 cell lines using Bayesian hierarchical effects model, *Proceedings of SPIE, BIOS 2001, Microarrays: Optical Technologies and Informatics,* Vol. 2, 23, 228-235.

Lee, J.K. (2002). Discovery and validation of microarray gene expression patterns, *LabMedica International*, 19, 2: 8-10.

Lee, J.K. and O'Connell, M. (2003) *An Splus library for the analysis of microarray data, The analysis of gene expression data: methods and software*, edited by Parmigiani, Garrett, Irizarry, and Zeger. Springer: New York.

Lin, J.K. and Thomas, D.C. (2000) Performance of Markov Chain Monte Carlo approaches for mapping genes in oligogenic models with an unknown number of loci. Am. J. Hum. Genet. 67: 1232-1250.

Nadon, R., Shi, P., Skandalis, A., Woody, E., Hubschle, H., Susko, E., Ramm, P. and Rghei, N. (2001) Statistical inference methods for gene expression arrays, Proceedings of SPIE, BIOS 2001, Microarrays, Optical Technologies and Informatics, 46-55.

Newton, M.A., Kendziorski, C.M., Richmond, C.S., Blattner, F.R., Tsui, K.W. (2001). On differential variability of expression ratios: Improving statistical inference about gene expression changes from microarray data . Journal of Computational Biology , 8 37-52.

Pinkel D (2000). Cancer cells, chemotherapy and gene clusters. Nat. Genet. Mar;24(3):208-9.

Scherf, U., Ross, D.T., Waltham, M., Smith, L.H., Lee, J.K., Kohn, K.W., Reinhold, W.C., Myers, T.G., Andrews, D.T., Scudiero, D.A., Eisen, M.B., Sausville, E.A., Pommier, Y., Botstein, D., Brown, P.O., and Weinstein, J.N. (2000). A cDNA microarray gene expression database for the molecular pharmacology of cancer. *Nature Genetics,* 24 (3), 236-244.

Stoeckert CJ, Causton HC, and Ball CA. (2002). Microarray databases : standards and ontologies, Nature Genetics, 32, 469-473.

Storey, J.D. and Tibshirani, T. (2003) SAM thresholding and false discovery rates for detecting differential gene expression in DNA microarrays, edited by Parmigiani, Garrett, Irizarry, and Zeger. Springer: New York. Tusher V., Tibshirani R. and Chu C. (2001). Significance analysis of microarrays applied to transcriptional responses to ionizing radiation, Proceedings of the National Academy of Sciences 98: 5116-21.

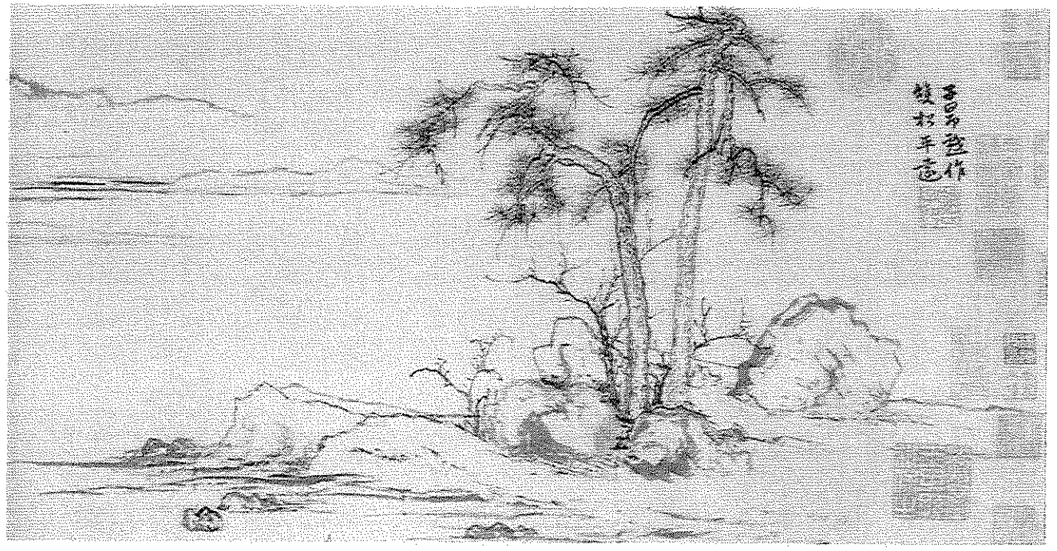
Wolfinger, R.D. and Gibson, G. and Wolfinger, E.D. and Bennett, L. and Hamadeh, H. and Bushel, P. and Afshari, C. and Pales, R.S. (2001). Assessing Gene Significance from c{DNA} Microarray Expression Data via Mixed Models, Journal of Computational Biology, 8, 37-52.

Jae K. Lee, Ph.D.
Division of Biostatistics and Epidemiology
University Virginia, PO Box 800717,
Charlottesville, VA 22908

Upcoming Controversial Statistical Issue (January 2005)

"Bioinformatics (part 2): Genomics (SNPs) Data"

Interested contributors please contact Dr. Sue-Jane Wang (wangs@cder.fda.gov).



Regional Activities

By: Mufa CHEN & Guo-ying LI

China Area

The Joint Meeting of CSPS/IMS

July 9-11, 2005, Beijing, China

http://math.bnu.edu.cn/statprob/CSPS-IMS2005/ index.html

The joint meeting of the Chinese Society of Probability and Statistics (CSPS) and the Institute of Mathematical Statistics (IMS) will take place in Beijing on July 9~12, 2005.

The venue of the meeting is Peking University with accommodations in the nearby Friendship Hotel (shuttles will be provided between the university and the hotel). As a city, Beijing offers many attractions both cultural and touristic. On behalf of the Program and Local Organizing Committees we are delighted to invite you to come to Beijing. Your participation will ensure that the 2005 CSPS/IMS joint meeting become an unforgettable scientific event.

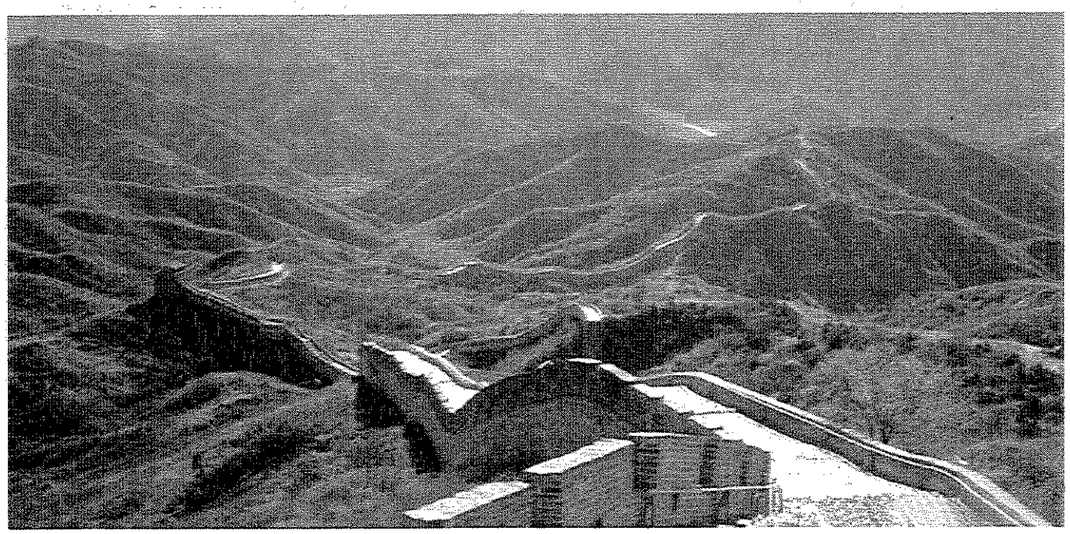
The invited program covers a wide range of topics in statistics and probability, presenting recent and state-of-the-art developments in modern methodology research and applications such as nonparametric statistics, machine learning, finance, bioinformatics, environmental statistics, and information technology.

*** Submissions of contributed papers are invited to the conference website with a deadline of Jan. 20, 2005. ***

Moreover, a half-day sightseeing to the Great Wall during the meeting is planned and an after-meeting program and an accompanying persons program during the meeting are also being planned. Please visit the conference website for updates.

We look forward to meeting you in Beijing!

Chairs of the CSPS Program Committee: Mufa CHEN & Guo-ying LI;
Chair of the IMS Program Committee: Bin YU;
Chairs of the Local Organizing Committee Zhi GENG & Shuyuan HE.



Regional Activities

Hailiang Yang, Ph.D.

Hong-Kong Area

Conference on Insurance Mathematics, Ruin Theory and Monte Carlo Methods, 28-30 June 2004:

This conference is jointly organized by the Department of Statistics and Actuarial Science and the Institute of Mathematical Research, Department of Mathematics, HKU, the captioned conference will be held from 28-30 June at the University of Hong Kong. The Conference will feature many leading experts and distinguished speakers, among whom are Hans Gerber, Marc Goovaerts, Tse-Leung Lai and Elias Shiu, etc. The Organizing Committee includes: T.L. Lai (Stanford and HKU), K.W. Ng (HKU, N.H. Chan (CUHK), Elias Shiu (U of Iowa and HKPU), H.Yang (HKU) and S.P. Yung (HKU). For more details, please go to the conference's website. The website address is http://hkustasc.hku.hk/users/irm/

Threshold models and new developments in time series

A conference entitled "Threshold models and new developments in time series', is being organized at the University of Hong Kong on July 12-14, 2004, in honor of Professor Howell Tong. The Organizing Committee includes: K. W. Ng, H.Z. An, K.S. Chan, N.H. Chan, T.L. Lai, W.K. Li, R.S. Tsay, Q.W. Yao. Speakers include many leading experts in time series and related areas. For details please go to the conference website at: http://hkustasc.hku.hk/users/ndts/

The Third International Congress of Chinese Mathematicians (ICCM 2004)

The Third International Congress of Chinese Mathematicians will be held in Hong Kong from December 17 to December 22, 2004. The Congress is sponsored by The Morningside Group. It is organized and sponsored by The Institute of Mathematical Science and the Department of Mathematics at The Chinese University of Hong Kong. It is supported by the sister universitites in Hong Kong and also the Mathematics centers at Academica Sinica and Zhejiang University. The Congress Chairman is Professor Shing-Tung Yau. For more details, please visit the conference website at: http://www.ims.cuhk.edu.hk/conference/iccm2004/

Hailiang Yang, Ph.D. is Associate Professor of the Statistics and Actuarial Science Department, The University of Hong Kong.
Email: hlyang@hkusua.hku.hk

Regional Activities

C. Andy Tsao, Ph.D.

Taiwan Area

The 2004 Southern Taiwan Statistical Conference and Chunghwa Data Mining Society Annual Meeting.
June 24—24, Agora Garden, Taipei.
http://stat.nccu.edu.tw/ssc2004/Frame.htm

Probability and Statistics Conference in honor of Dr. Chow, Y.S.'s 80th Birthday.
June 26. Academia Sinica, Taipei.
http://www.stat.sinica.edu.tw/yschow/schedule.htm

2004 Statistics Camp for Undergrads
June 19—20.
Academia Sinica, Taipei.
http://www.stat.sinica.edu.tw/camp2004/

2nd Intelligence Technology and Applied Statistics Conference
June 19—20.
National Chung-Cheng University, Chiayi.
http://www.math.ccu.edu.tw/chinese/news/2004/news0608.htm

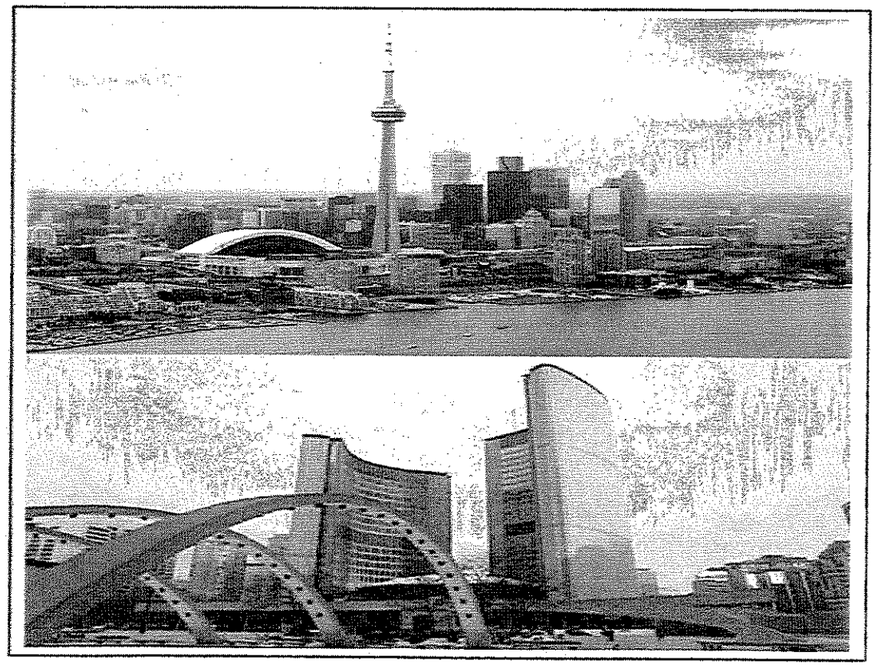
C. Andy Tsao, Ph.D. is an Associate Professor of Statistics, Department of Applied Math at National Dong Hwa University, Taiwan.
Email: chtsao@mail.ndhu.edu.tw.

ICSA Annual Meeting in Toronto

By: Annual Meeting Committee

On behalf of the annual meeting committee, welcome you to Toronto! Toronto is the home to the world's tallest building (CN Tower at 553.33m) and that the world's longest street (Yonge Street). Toronto is one of the most multicultural cities in the world and is ranked as the safest large metropolitan area in North America. As our tradition, ICSA will have a booth and will hold its annual members meeting during JSM. If you are planning to attend the JSM, please come visit the ICSA booth to find out what is new at ICSA and to meet old or new friends. This year, our annual members meeting is from 6:00 to 7:00pm of Wednesday, August 11 at Room 707 of the Toronto Convention Center. Following the meeting, we will have a dinner (at 7:30pm) in a famous authentic Cantonese restaurant (Dynasty) at 131 Bloor St. West. It is within 25 minutes walking distance, and can be reached by subway conveniently from the most hotels for JSM. Tickets can be purchased at ICSA booth in JSM and cost $40cdn per person.

Toronto has many local attractions. The City Hall with its distinctive design has won numerous architectural awards for its curved twin towers and the central pod. A list from the city website includes Casa Loma (Toronto's famous castle), Entertainment District (pay Canadian Dollar for the North American Standard shows), Ontario Place, Ontario Science Centre, Toronto Metro Zoo, and Centreville Amusement Park (Central Islands). You will surely visit China town nearby conference hotels.



ICSA 2005 APPLIED STATISTICS SYMPOSIUM

June 19-22, 2005 at Washington, DC, U.S.A.

The ICSA Applied Statistics Symposium Program Committee would like to invite your participation of the 14th annual ICSA Applied Statistics Symposium to be held at Washington DC Metropolitan Area

DATE: June 19 to 22, 2005.

PROGRAMS: Short courses will be held on June 19 and technical sessions will be held on June 20 to June 22.

LOCATION AND ACOMMODATION: To be announced.

CALL FOR PAPERS: The Program Committee invites you to submit statistical papers to be considered for presentation at the symposium. Abstract for the contributed papers are due on February 28, 2005.

STUDENT AWARD AND TRAVEL FELLOWSHIPS: The Program Committee will grant several travel fellowships to student with excellent papers.

J.P. HSU MEMORIAL STUDENT AWARD: The Program Committee will grant one J.P. Hsu Memorial travel fellowship for student paper on medical statistics.

Details of submission and template of abstract and student paper can be found at the ICSA website http://www.icsa.org. The tentative program will consist of 5-6 short courses and about 30 invited sessions. For further information, please visit ICSA website.

2005 Applied Statistics Symposium Committees:

Ling Chen (Chair of Short Courses Committee), chenli@cder.fda.gov
Jinbo Chen, chenjin@mail.nih.gov
Milton C.L. Fan (Chair of Registration Committee), fanm@cder.fda.gov
Shibao Feng, sf96@georgetown.edu, Ping Hu, pingh@mail.nih.gov
James Hsien-Ming Hung (Chair of Program Committee), hung@cder.fda.gov
Jia-wen Ko, cko5@jhmi.edu, Chenxiong Le, lec@cder.fda.gov
Hung-Ir Li (Chair of Fund Raising Committee), Li_Hung-Ir@Allergan.com
ShuHua Li (Chair of Banquet Committee), sli1@nida.nih.gov
Zhao Hai Li, (Chair of Student Helpers Committee), zli@research.circ.gwu.edu
Aiyi Liu, liua@mail.nih.gov, Chih C. Stan Lin, linst@cder.fda.gov, Ming Tan, mtan@umm.edu
Yi Tsong (Chair of General Program Committee), tsong@cder.fda.gov
Sue-Jan Wang, wangs@cder.fda.gov, Colin O. Wu, wuc@nhlbi.nih.gov
Lap-Ming Wun (Chair of Logistic Committee), lwun@AHRQ.gov
Grace Yang (Chair of Student Award Committee), gly@math.umd.edu
Gang Zheng, (Chair of 2005 Symposium website Information Committee), zhengg@nhlbi.nih.gov

International Chinese Statistical Association

泛華統計協會

Membership Application & Renewal Form

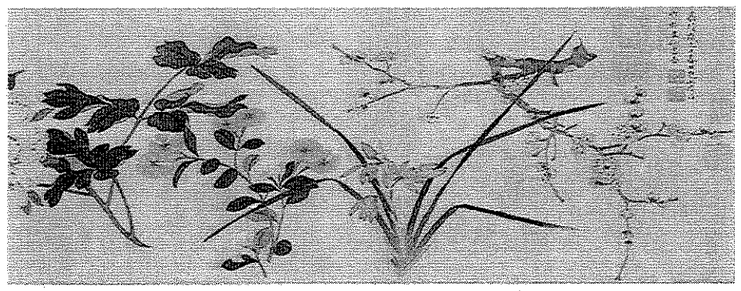
Name	(Last)	(Middle)	(First)
(English)			
(Chinese)			
Address			
Office	Address:		
	City:		
	State:	Zip Code:	Country:
	Email:	Telephone:	FAX:
Home	Address:		
	City:		
	State:	Zip Code:	Country:
	Email:	Telephone:	FAX:
Education			
	Degree:	Year Graduated:	
	University:		
Professional Occupation & Title			
	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, 105 Oval Lane, North Wales, PA 19454			

International Chinese Statistical Association

Profit and Loss

January 1, 2004 through June 30, 2004

Ordinary Income/Expense	
Income	
Advertisement	300.00
Donations	1,070.00
Miscellaneous	52.59
Membership Dues	2560.00
Total Income	**3982.59**
Expense	
Miscellaneous	
Member service	*94.98*
Working lunches	*67.40*
Refund to symposium account	*529.50*
Total Miscellaneous	691.88
Postage and Delivery	
Account documents	*345.27*
Book/Journal Donation	*212.00*
Bulletin	*1475.71*
Total Postage and Delivery	2032.98
Printing and Reproduction	
Jan. Bulletin	*3760.00*
Total Printing and Reproduction	3760.00
Web Page Hosting	1754.29
Total Expense	**8239.15**
Net Ordinary Income	-4256.56
Other Income/Expense	0
Net Other Income	0
Net Income	-4256.36

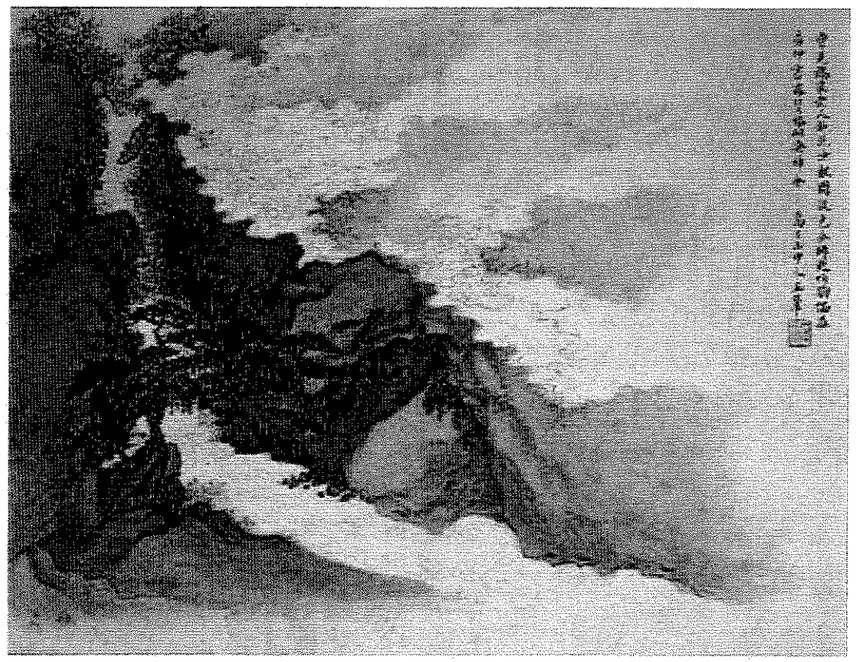


International Chinese Statistical Association

Balance Sheet

January 1, 2004 through June 30, 2004

ASSETS	
Checking/Savings	
Checking	60029.85
Savings-Money Market	0
TOTAL ASSETS	**60029.85**
LIABILITIES & EQUITY	
Equity	
Opening Balance 1/1/2004	64,286.41
Net Income	-4256.56
Total Equity	60029.85
TOTAL LIABILITIES & EQUITY	**60029.85**



Calendar of Meetings

July 21-23, 2004 - The Sixth ICSA International Conference
Location: Singapore.
More information available at: Zhiliang Ying zying@stat.columbia.edu.

July 26-30, 2004, IMS Annual Meeting/ 6th Bernoulli World Congress
Location: Barcelona, Spain
http://www.imub.ub.es/events/wc2004/

The program covers a wide range of topics in statistics and probability, presenting recent developments and the state of the art in a variety of modern research topics and in applications such as mathematical finance and statistical bioinformatics. The program include up to twelve Special Invited Lectures given by leading specialists, thirty-five Invited Paper Sessions and a large number of contributed talks. David Nualart is the Chairman of the Organizing Committee and Wilfrid Kendall is the Chairman of the Scientific Committee.

August 4-6, 2004 The Seventh North American New Researchers Conference
Location: York University, Toronto, Canada

The New Researchers' Conference (NRC) is an annual conference sponsored by the Institute of Mathematical Statistics dedicated to fostering interaction between new researchers in the fields of Probability and Statistics. The Seventh NRC will be held at York University in Toronto, August 4-7, 2004, immediately prior to the Joint Statistical Meetings, also to be held in Toronto

The purpose of the conference is to provide a comfortable setting for new researchers to share their research and make connections with their peers in an informal setting. The conference is kept relatively small (50-60 participants) so as to maximize the intellectual and social interaction. All participants are expected to give a short, expository talk or contribute a poster on their research. Anyone who has received a Ph.D. since 1999 is eligible to attend, although priority will be given to first time participants. Abstracts of these talks and posters will appear in the IMS Bulletin.

Deadline for receipt of applications is February 15, 2004. Please apply promptly since the number of participants is limited. In addition, we expect to receive funding to defray travel and housing costs for participants. To apply, please submit letter of interest curriculum vitae, and title and abstract of presentation preference for a talk or poster to

Peter Song
Dept of Mathematics and Statistics
N520 Ross Building, 4700 Keele St
York University, Toronto, Ontario, Canada M3J 1P3
Email: song@mathstat.yorku.ca
http://www.math.yorku.ca/StatsSection/NRC

August 6-7, 2004, New Directions in Probability Theory
Location: Fields Institute, Toronto, Canada
http://www.imstat.org/meetings/NDPT/default.htm

The meeting is co-sponsored by the Institute of Mathematical Statistics (IMS) and the Fields Institute for Research in Mathematical Sciences. It will be held at the Fields Institute.

It is intended for a general probability audience interested in recent developments in probability theory. The topics of the session are Random Walks with Self-Repulsion, Random Matrices, Random Media, Super-processes, and Markov Chains with Algorithms. There will be no registration fee for the meeting. However, space at the Fields Institute is limited and early registration is recommended.

January 12-14, 2005 The Second Joint IMS/ISBA International Conference
Location: Bormio, Italy (Italian Alps)
http://www.eco.uninsubria.it/IMS-ISBA-05/

July 9-11, 2005 The Joint Meeting of the Chinese Society of Probability and Statistics CSPS) and the Institute of Mathematical Statistics (IMS)
Location: Beijing, China
http://math.bnu.edu.cn/statprob/CSPS-IMS2005/index.html/
(Details on page 52)

Our Sincere Thanks!

The Editorial Team

As always, many good friends have taken time from their busy schedules to write for the Bulletin. Without their help, this issue of the Bulletin would not be possible.

We especially appreciate the generosity of Dr. Kettenring to take time for us to conduct the interview and to review and revise the draft in lightning speed. His kindness made our life seem so easy.

It is remarkable to have had Dr. Peace join us to address the Jiann-Ping Hsu's memorial session and established the endowment. His generosity will be appreciated and cherished forever by all of us.

We would also like to express our thanks towards Professors Mufa Chen and Guo-ying Li, Dr. Hailiang Yang, and Professor Tsao for providing us with the local statistical activities in their respective regions. These efforts will greatly enhance the interactions between the statisticians around the world.

Of course, the Symposium Committee also deserved our loudest applause not only for their dedication to the successful meeting but also for their contributions of the meeting report and lots of wonderful photos of highlights. Due to the limitation of space, we cannot publish all of the beautiful pictures. Fortunately, we can visit the ICSA website for the whole gallery of pictures.

The controversial statistical issues continue to be one of our highlights in the Bulletin. The contributors indeed deserved our thanks for these highly educational viewpoints.

As usual, we would like to thank the editorial team for their time and dedication to put together these articles and publish them in time for the readers.

We hope to have more helping hands from our friends and members for the future issues of the ICSA Bulletin.