International Chinese Statistical Association

http://www.icsa.org

泛華統計協會

Features:

Monte Carlo & Bioinformatics

Interview with a Distinguished Statistician

Statistical Issue: Flexible Design

Meeting Announcements

Bulletin January 2003

Mao-kung Ting
(Late Western Chou Dynasty, ca.9th~8th centuries B.C.)

The Mao-kung Ting is a 2800-year-old ritual bronze vessel that was used at the ancestral temple of the court. The 500 characters cast on its inside surface form the longest inscription found on the thousands of bronzes remaining from the Shang and Chou periods. King Hsuan Chou's uncle, Duke of Mao, ordered the casting of this precious ting in gratitude for the King's generosity to his fellow citizens and as a memorial for his heirs. The Mao-kung ting is among the world's most invaluable treasures. It is now part of the National Palace Museum collection.

From the Editor

Kao-Tai Tsai, Ph.D.

To be the Editor of the ICSA's Bulletin is not an easy decision. It takes someone with a moment of foggy mind to make the commitment because as in old Chinese saying, "To create something new is difficult, however, it is not easy either to maintain the existing level of success".

This is a perfect reflection of the production of this issue of the Bulletin. With the current stage of success from the great effort of the previous Editors, it is really a challenge for a new Editor to take over the post and to maintain the high level of quality with the transition of all the ICSA committees including the President. However, with a dedicated Editorial Team, we somehow managed to pull things together in a hurry within our spare time.

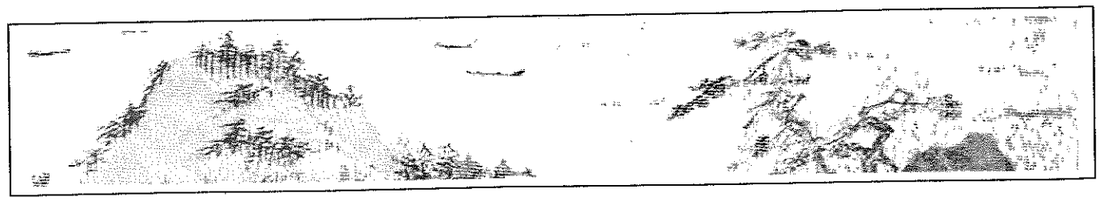
In this issue, in addition to maintaining the previous works, we also spice it up with some new thinking. We change the appearance of the Bulletin to reflect something more Chinese. It is good to be a great statistician, however, it is so much better to enrich ones life with a good appreciation of wonderful Chinese culture. For this purpose, we included some classical Chinese masterpieces such as great paintings and calligraphy by famous artists. Ideally, we would like to have them all in bright colors, however due to budget considerations, we can only do it black and white in this issue. Hopefully, we can improve this in the near future.

We also reserve a page of Letter-to-the-Editor for you to voice opinions. Please tell us your thoughts so that we will know whether the Bulletin is serving its purpose. I sincerely urge everyone to pick up your pen and write to us, contribute your articles and your thoughts.

Just as in this Chi Baishi's painting, the Bulletin is ours and it is so much fun to ***JOIN THE CLUB!***

Table of Contents
ICSA Bulletin, January 2003

Editorial Members
ICSA Bulletin, January 2003



Committee Chairs
ICSA, 2003

Board of Directors

Ngai Hang Chan	William W.S. Wei
Chen-Hsin Chen	Zhiliang Ying
Rong Chen	Heping Zhang
Mei-Ling Lee	Frank Shen
Guo-Ying Li	Jiahua Chen
Zhaohai Li	Greg Wei
Don Sun	Ivan Chan
Naitee Ting	Shu-Yen Ho
Sue-Jane Wang	Xumin He

Standing Committee
Program Committee:
Wei-Yann Tsai.
Finance Committee:
H.M. James Hung
Nominating & Election Committee:
Naitee Ting
Publication Committee:
James J. Chen.
Constitution Committee:
Frank Shen.

Current Committee
Membership Committee:
Tzu-Cheg Kao.
Fundraising Committee:
Alice Hsuan
Public Relations Committee:
Yi Tsong.
Awards Committee:
Lynn Kuo
Communication Committee:
Rong Chen
Applied Statistics Symposium Committee:
William W. S. Wei
Book & Journal Donation Committee:
Tar Timothy Chen
Annual Meeting Committee:
Wei-Yann Tsai
Archive Committee:
Yi Tsong
Strategic Committee:
Chao Agnes Hsiung

Submission Guidelines
ICSA Bulletin

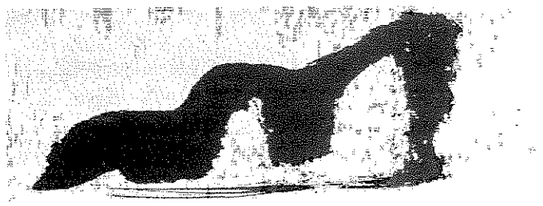
Articles
The ICSA Bulletin welcomes the submission of articles by the members. Articles submitted should be written in traditional Chinese or English. The contents should be aimed at the general reading.

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

This space is reserved for your opinions and your creative ideas. Please send in your thoughts, your suggestions, to let us know the things you like (the good), the things you dislike (the bad), and the things you absolutely do not want to see (the ugly).

This space belongs to every member and I hope we can fill this space up for the next issue.

International Chinese Statistical Association

泛華統計協會

From the President

Zhiliang Ying, Ph.D.

Dear Fellow ICSA Members:

As we leave the year of the horse, ICSA has experienced another year of success under the leadership of Professor William Wei and Dr. Yi Tsong. The foundation of the organization has been further solidified with an updated Constitution, thanks to a hardworking team led by Frank Shen. The standing and current committees have been reorganized to better reflect their respective functions and to improve the organization's efficiency.

Although it is a U.S. registered nonprofit organization, ICSA has a wide international appeal with its members coming from all over the world. For years we have been working hard on emphasizing this global nature and striving to establish a broad membership base, particularly in the East Asia region. Thus far, five ICSA conferences have been held in, chronologically, Hong Kong, Taipei, Beijing, Kunming and Hong Kong again. We are delighted that our colleagues in Singapore, led by Professor Louis Chen, are organizing the Sixth ICSA Conference, to be held in the summer of 2004. This will be the first time we sponsor a conference in Singapore, an internationally oriented country with a diverse culture. Please look for the detailed announcement in this issue of the Bulletin. The ICSA website (http://icsa.org/) will soon have updated information about the conference as well as relevant events and attractions. We encourage our members to take the opportunity to attend the conference.

Another event in the East Asia region is the Fourth International Conference on Multiple Comparisons, for which ICSA is a co-sponsor. The venue and dates are to be decided in order to coordinate with the ICSA Conference in Singapore so that members and interested parties can attend both conferences. This is the kind of collaboration ICSA seeks with other professional organizations. For inquiries regarding ICSA conference co-sponsorship, please contact Dr. Naitee Ting, Chair of the Program Committee, or myself.

The next upcoming major event for us is the ICSA 2003 Applied Statistics Symposium in San Diego, California. It is the second time that this annual event ventures outside the Northeast Corridor of Boston to Washington D.C. The last and only time that it was held outside the Northeast Corridor was the 2001 Symposium in Chicago. I am sure that the mild southern California climate with its hospitality and beautiful coast will attract a large attendance. The shorter traveling time from the East Asia region provides an additional incentive for our colleagues from that part of the globe. Remember, suggestions and proposals for future symposium sites are always welcome.

We were fortunate to have Dr. Sue-Jane Wang as the Editor-in-Chief of the ICSA Bulletin for the past three years. She did a super job in expanding the scope of the Bulletin and making it such an attraction. We are also fortunate that Dr. Kao-Tai Tsai was willing to take the job as the new Editor-in-Chief. This will entail many sleepless weekends and holidays for him in the coming years. I know that Kao-Tai already started working on this issue of the Bulletin months ago by collecting materials and contacting people all over the world. I am sure that under his leadership, the Bulletin will continue to flourish.

An important development, which Dr. Don Sun, the web master and current Chair of the Communication Committee, is working on is to establish an online balloting system. The online membership directory access and update of membership information are being worked upon as well.

The ICSA Constitution and By-Laws stipulate that we elect a new president every year. New members for the Board of

From The President...

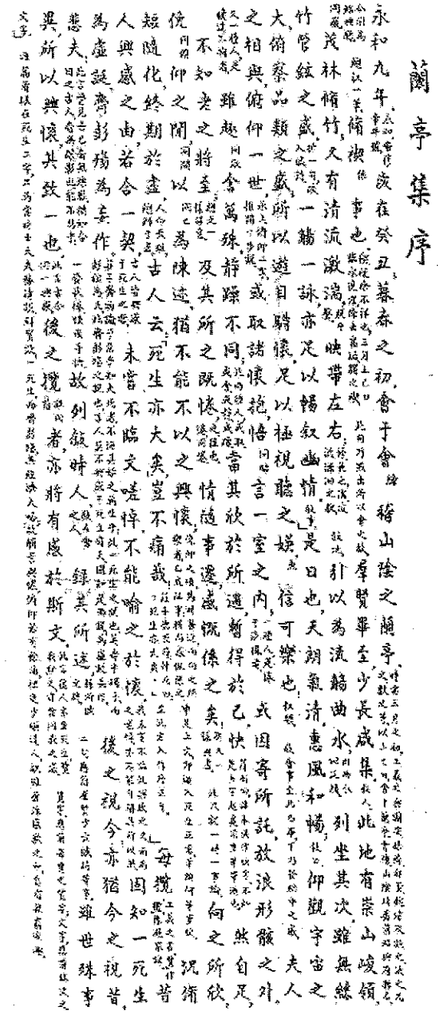
Directors also need to be elected. At the last Joint Statistical Meetings, Frank Shen was elected as the President along with a number of new board members. Congratulations to all of them. For suggestions regarding potential candidates, please contact the Nominating and Election Committee. It is important that members cast their ballots in the upcoming election.

Any professional organization needs a constant influx of new members to sustain its growth. ICSA is no exception. We need to be vigilant about recruiting new members. Statistics show that new members and therefore growth will largely come from current and future graduate students and from recruitment in industry and in the East Asia region, where the ICSA membership is currently relatively low. To this end, we must overcome gaps between younger and older generations, between industry and academia, and between East Asia and North America. We hope that through the efforts of our members and through activities such as the upcoming conferences, there will be more awareness of our organization and, consequently, a broader membership.

Looking ahead into 2003 and beyond, we have many important activities under preparation. We also face many challenging tasks. I am optimistic that through our collective wisdom, efforts and dedication, ICSA will continue to grow. Finally, I wish all of you a happy, successful and productive new year.

Zhiliang Ying, President, ICSA

Zhiliang Ying, Ph.D., is a Professor of the Statistics Department, Columbia University, New York, New York, USA.


蘭亭集序

Calligraphy above: Preface of Lanting by Wang Hsi-chih, (303-361)

This is the famous "Preface of Lanting" composed and written in 353 AD by the great Chinese calligrapher, Wang Hsi-chih, to commemorate a wonderful gathering of scholars, intellectuals, and, most important, good friends.

From the Past President

William W. S. Wei, Ph.D.

My Reflections on the Year 2002

It has been a real honor and pleasure to serve ICSA as its 2002 President. The Editor-in-Chief of the Bulletin, Dr. Kao-Tai Tsai, asked me to write an article with my reflections as ICSA President and share with our members what we have accomplished and what still needs to be improved upon.

Many members have told me that a one-year term is too short to make any significant impact, but I disagree. In fact, I actually started working in 2001. As the 2001 President-elect, my first responsibility was to refresh myself on the ICSA Constitution and By-Laws, the recent Board of Directors meeting minutes, and important events described in recent issues of the Bulletin so that I could fully understand my responsibilities.

The spirit of the Constitution clearly implies that the Presidency is not a privilege but a responsibility. It is not simply a one-person job. Instead, it is about teamwork. The team includes the President, Board of Directors, the Executive Director, the Treasurer, as well as the chairs and members of various committees. An important role of the President is to form a unified, hard-working team in addition to providing leadership for both the team and the Association. With these concepts in mind, I started to search for competent and willing members to chair various committees after the 2001 Joint Statistical Meetings. Traditionally, board members read proposals written by the various committees. In order to help board members become more familiar with the inner-workings of these committees and thus the Association as a whole, I began my search among members of the board. In the process, I found that the most important tasks were primarily done by temporarily assigned current committees instead of standing committees. This deeply troubled me, and I was extremely concerned about the potentially undesirable long-term consequences. Later, this worry became one of the driving forces for the team and myself in 2002. Some important changes and improvements are listed below. Included in the list are both completed projects as well as those that have already begun but are currently still on-going.

(1) Following the By-Laws and through the effort and cooperation of the Constitution Committee and Board of Directors, we successfully revised the Constitution. As a result of the revision, the Program Committee, the only standing committee of ICSA that deals with meetings, will no longer simply organize the banquet at Joint Statistical Meetings. Instead, in addition to assisting the newly established Annual Meeting Committee in planning the annual meeting, it will recommend symposium and conference sites, including candidates for their chairs, and more importantly, it will recommend general policy for all meetings. Similarly, the Finance Committee will no longer only deal with income, expenditures, and the budget. Instead, it will have the added responsibility of long-term financial planning, including the investment of the Association's assets.

(2) The Association holds board meetings and membership meetings twice a year – once at the ICSA Applied Statistics Symposium and again at the Joint Statistical Meeting. I was charged with arranging the 2002 Applied Statistics Symposium in 2000 before becoming the President-elect. Thus, organizing the Symposium also became one of my responsibilities in 2002. I needed to form

From The Past President...

another team to help make the event a success. Because of the dedication and sacrifices of the 2002 Symposium Program Committee members, the Symposium held in greater Philadelphia between June 6 and 8, was a great success after two years of planning. A record-breaking of two hundred sixty-two participants joined the Symposium, and more than one hundred members attended the membership meeting on June 7. Through successful fundraising and a record number of attendees, the Symposium had a record-breaking surplus of $30,000. In addition, at the board meeting on June 5, we discussed many important issues and began working on them the next day.

(3) We formed a new E-Directory Committee to conduct a survey and study the feasibility of an e-directory. This should become a reality soon.

(4) Because of new technology, e-mail has become the most common and efficient communication tool. Many organizations have stopped using paper communication entirely and rely solely on e-mail for communication and announcements. This includes conference announcements, registration, and even elections. It is important for the Association to keep the most updated e-mail system possible. With regard to elections, the voting rate of our members was not very good. To improve the process, the Nominating and Election Committee has been studying the possibility of electronic voting. The proposal has been submitted, and we hope that it will be adopted and improve the system next year. The success of this project requires an updated membership directory and a working e-mail system, which needs the on-going effort of the Executive Director and the Membership Committee.

(5) As a professional organization, it is important for us to have a systematic means of storing and saving important records for our Association. At the June 7 board meeting, the issue of an electronic archive was raised and discussed. The work is clearly extensive and needs to be consistent and continuous. In addition, those in charge of this task should be knowledgeable about the Association. Thus, the Board approved an Archive Committee consisting of the Association's former and present Executive Directors with the current Executive Director as its chair. The Executive Director will continue to report the progress of the project to the board.

(6) To serve our members and profession, ICSA organizes a yearly Applied Statistics Symposium in the U.S. and tri-yearly International Conference in Asia. The 2001 International Conference was held in Hong Kong, and its announcement appeared in every issue of the Bulletin since July 1999. However, no announcement for the 2004 ICSA International Conference was made in the Bulletin this year. Since attending an international conference normally requires more time for planning and budgeting, it is important for our members to be made aware of the event at least one year ahead of time. In order to finalize the conference dates, I went to Singapore in July and settled them with Prof. Louis Chen and his colleagues at the National University of Singapore. The Conference will be held in Singapore between July 21 and 23, 2004. The conference announcement will appear in the January 2003 issue of the ICSA Bulletin.

(7) We continued to work with various organizations through communication and co-sponsorship of meetings and events. These organizations include professional statistical organizations, companies, universities, and research institutes. We have been invited again to submit a proposal for an invited session at the 2003 Joint Statistical Meetings. The proposal has been submitted and accepted for the 2003 JSM.

From The Past President...

(8) We continued to support a book/journal donation service. We first began providing the service to universities in China, but now through the effort of the Book and Journal Donation Committee, the service has been extended to other countries such as Pakistan.

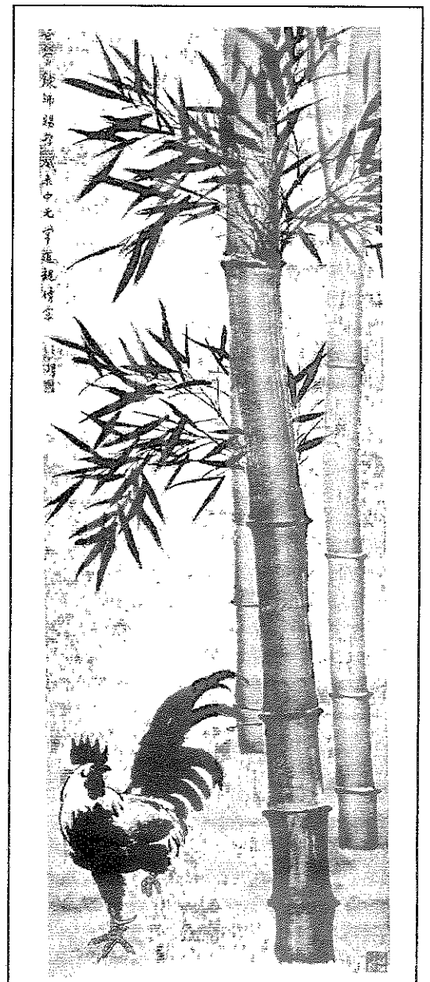
(9) Dr. Sue-Jane Wang completed her exceptional three-year term as the Bulletin's Editor-in-Chief at the end of 2002. The Publication Committee has selected Dr. Kao-Tai Tsai to succeed her from 2003 to 2005.

(10) The second membership meeting was held on August 14, 2002, at the Joint Statistical Meetings in New York. Although there were not as many members present at that membership meeting as there were at the June meeting in Greater Philadelphia, we still had some good discussions. The dinner was held at the Peking Park Restaurant, and the turnout was excellent. We thank the Annual Meeting Committee for helping to organize these successful events, and the many graduate students from Columbia University for staffing the help desk during the week-long conference in New York.

Thank you for your help and support. I would also like to thank the Directors of the Board, chairs and members of the committees, the Executive Director, and the Treasurer, for their service and dedication to the Association. I know you will all join me in giving our support to the next President, Professor Zhiliang Ying.

Have a Happy New Year and a wonderful holiday.

William W.S. Wei, Ph.D., is a Professor of the Statistics Department, Temple University, Philadelphia, Pennsylvania, USA.



Painting Above: A Reflection by Master artist Beihong Xu.

Xu specialized in Chinese paintings of birds, flowers, and bamboos. He is especially well known for his paintings of horses. All of his paintings emphasized the capturing of the spirit of the objects in the pictures to make them vividly alive.

From the Executive Director, ICSA

Yi Tsong, Ph.D.

Dear friends:

Happy New Year! With the effort of our Ex-President William Wei, Officers, Board of Directors, Committees and support from all of you, our organization enjoyed a very successful year in 2002. Now, we are opening the New Year 2003 under the new management crew including President Zhiliang Ying and Editor-in-Chief of ICSA Bulletin Kao-Tai Tsai. We regret to see the departure of our President William Wei and Editor-in-Chief of ICSA Bulletin, Sue Jane Wang. Under his direction, restructure of the Committee was made for more efficient operations. Frank Shen and his Committee revised the ICSA Constitution and By-laws. During her 3-year tour of duty, Sue Jane and her editorial team transformed the ICSA Bulletin into a lively journal. However, an old solider will never die and she will soon become the Chair of the Biometrics Section in 2004. On the other hand, as a member of Sue Jane's editorial team, it is for sure that Kao-Tai had learned many tricks. He will certainly open your eyes even wider. You just wait and see.

In addition, many Directors of the Board and Chairs of Committees and Biometrics Section are also retiring from their position to assume new duties. Retiring this year are Jianqing Fan, Agnes Hsiung, Ker-Chau Li, Kar K. Lin, Jun Shao, and Mei-Cheng Wang from the Board; Frank Shen from representative of Biometrics Section, James HM Hung from Chair of Biometrics Section. We can never thank them enough for their unselfish contribution and endless energy invested in their works.

Luckily, we found many members in our Association who are so eager to make significant contribution in the future of this Association. Let us welcome the new officers and Directors elected last June: 2004 President-elected - Dr. Frank Shen; 2003 Board of Directors - Drs. Jiahua Chen, Greg Wei, Ivan Chan, Shu-Yen Ho, and Xumin He;
2004 Chair-elected of Biometrics Section – Dr. Sue Jane Wang;
and the new Chairs appointed by President Zhiliang Ying.

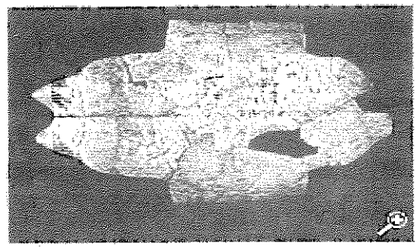
The membership directory is in print now and will be delivered to your mailbox in this month. Hopefully, it is a useful document to help you locate and communicate with your colleagues across the land and seas.

Please remember to take time to visit the ICSA web site frequently. You will find much timely and important information posted there.

Hope you have a Grand New Year!

Sincerely yours,
Yi Tsong

Yi Tsong, Ph.D. is a Mathematical Statistician at the US Food and Drug Administration, Rockville, MD., USA.
Telephone: (301) 827-3206.
Fax: (301) 480-2825.
Email: tsong@cder.fda.gov.



Picture Above: Turtle shell used to record ancient Chinese.

ICSA Membership Meeting Minutes – Aug. 2002

Yi Tsong, Ph.D.

Date: August 14, 2002
Time: 6:00 PM - 6:45 PM
Place: Hilton Hotel, New York City
Chair: William Wei (President)
Minutes: Yi Tsong
Attendees: About 70 ICSA Members

1. President Report

Prof. William Wei expressed gratitude to the local committee chaired by Prof. Wei-Yann Tsai for making the banquet arrangements and providing services at the ICSA help desks. He congratulated the following ICSA members who received honors and awards at the Joint Statistical Meetings:
(1) Prof. Jun Liu received the 2002 COPSS Award,
(2) Prof. Rong Chen and William Wei were elected as new ASA Fellows,
(3) Prof. Xihong Lin received Second Noether Young Scholar Award.

He also congratulated the success of members who were elected to various ASA offices. They include Professors Joseph Gastworth, Dennis Lin and Xiao-Hua Zhou.

He reminded members on the 2003 Applied Statistics Symposium to be held in San Diego. He also announced that the 2004 ICSA International Statistical Conference will be held in Singapore between July 21 and 23, 2004 and encouraged members to attend these conferences.

In closing, he urged members to visit ICSA website to update their e-mail addresses in order to improve the communication between the Association and its members.

2. Honor Ceremony

President William Wei announced the three winners of ICSA Distinguished Service Award nominated by the Award Committee. The 2002 awardees are:

Agnes C Hsiung, Ph.D.
Citation of the Award - For promoting biostatistics and genomic statistics and for leadership as the President of ICSA and the Director of the Division of Biostatistics and Bioinformatics of the Taiwan National Health Research Institutes. For various other contributions to ICSA including policymaking and long-term planning while serving on the Board of Directors and the Strategic Committee, and editorial assistance while serving as an Associate Editor of *Statistica Sinica.*

Ker-Chau Li, Ph.D.
Citation of the award - For leadership in promoting statistics in genomic research, for excellent and dedicated editorial service for the *Statistica Sinica*, for improving the quality of *Statistica Sinica* and shortening its reviewing times while serving as one of its Chair Editors. For other contributions including serving on the Board of Directors of ICSA and as an Associate Editor for *Statistica Sinica*

Yi-Ching Yao, Ph.D.
Citation of the Award - For leadership in promoting statistics in genomic and financial researches, for excellent and dedicated editorial service for the *Statistica Sinica*, for improving the quality of *Statistica Sinica* and shortening its reviewing times while serving as one of its Chair Editors. For other contributions including serving as an Associate Editor for the *Statistica Sinica*

3. Executive Director Report

Yi Tsong announced the results of the 2002 election of officials and Board Directors on the behalf of the Nomination Committee. The results are:

2004 President-elected – Dr. Frank Shen,
2003 Board of Directors – Drs. Jiahua Chen, Greg Wei, Ivan Chan, Shu-Yen Ho and Xumin He,
2003 Chair-elected of Biometrics Section – Dr. Sue Jane Wang.

4. Announcement of future ICSA Applied Statistics Symposiums

Dr. Nancy Lo announced that the 2003 ICSA Applied Statistics Symposium is to be held at the University of San Diego, California with the preliminary program to be published in the January issue of the ICSA Bulletin

Dr. Yi Tsong also announced that the Board of Directors approved that the 2004 ICSA Applied Statistics Symposium to be held at Washington D.C. Metropolitan Area.

Meeting adjourned at 6:45 PM for Banquet.

Information Update

Following members please send your current address & email to Yi. Tsong at tsong@ cder.fda.gov

Bob An	Tan Au
Chen Yang	Hui-May Chu
Irwin Ho	Bi-Min Hsu
Chun Gao	Jin Zhu
Wen-Yao Ku	Yungtai Lo
Xianggui Qu	Winston Taamq
Chaiho Wang	Jingdong Xie
Xujie Yu	Daozhi Zhang
George C. Chao	Liang-Shi Chen
Yufen Carol Chung	Won-Chin Huang
Peter Hu	Cheng Cheng Hu
Yufei Huang	Chia-Wen Ko
Kok Lip Ng	Peng Roger Qu
Jason J. Tian	Albert T. Wang
Li-An Xu	Nehemiah Cherng
Jerry Zhang	Yufen Carol Chung

From the Chair, Biometrics Section

Jen-pei Liu, Ph.D.

A Changing World —

As we move very rapidly into the third year of the new millennium, great achievements have been accomplished in biology and medicine. For example, the Human Genome Project has completed sequencing the entire human genome and many breakthrough drugs have been developed for treatment of previously incurable/untreatable diseases. Despite these great successes, more challenges are ahead of us. AIDS is more epidemic in most populous regions of the world such as sub-Sahara area of Africa and Southeast Asia. However, effective therapy for treatment and slowing epidemic of AIDS seem so near but in fact are so far away. Other challenges include recurrence of some old diseases such as tuberculosis in epidemic proportions, world famines, increasing cost for development of new drugs, etc.

As Biostatisticians or Biometricians, we have involved with and contributed significantly to the great success in biology, medicine and health in the past, we must continue to commit ourselves to face these future challenge to improve the world a healthier place. As in the context of Biometrics Section of the International Chinese Statistical Association, how can we do to achieve the above goals? I think that we must know ourselves, we must know each other, we must educate ourselves, and we must reach out. Therefore, as the new Chair of Biometrics Section, I would try to use the following approaches to serve the members of Biometric Section:

(a) Conduct a survey for all members of Biometrics Section. This survey allows us to know ourselves better. From this survey, the demographics, education level, geographical distributions and profession profiles of our members can be provided. Most importantly, needs of our members of Biometrics Section will be heard from the survey.

(b) Establish a Biometrics Forum at the ICSA website. Biometrics Forum is a place for our members to communicate each other to known each other better. The articles for the Forum can be a review of statistical designs and methods for a particular area, or they can be opinions on certain issues. They do not have to be academic research. They can be experience sharing or question posing. In summary, the Biometrics Forum will serve as a vehicle for our members to know each other.

(c) Continue to support and coordinate the short courses at the ICSA Applied Statistics Symposium. However, we must understand the future is in the next generations of students. Through the survey, Biometrics Section will know the status and needs of student members. Not only will we improve recruiting students to become our members but also we will actively recruit students into Biostatistics or Biometrics.

(d) Encourage our members to actively participate in local communities or other non-statistical professional societies to show what we can do for the society.

The proposal will be implemented after they are discussed and your valuable comments are incorporated and are approved by the ICSA Board of Directors. Please send your comments to me. With your help, Biometrics Section will move a tiny step toward improving this changing world into a healthier, cleaner, and safer place to live.

Jen-Pei Liu, Ph.D. is Professor of the Statistics Department, National Cheng-Kung University, Tainan, Taiwan.
Email: jpliu@email.stat.ncku.edu.tw

From the Past Editors, Statistica Sinica

Ker-Chau Li, Ph.D. & Yi-Ching Yao, Ph.D.

First, we truly appreciate the outstanding service of our Associate Editors and referees. During the past three years of our term (Aug 1, 1999 - July 31, 2002), we received 539 new submissions, a growth of about 20 percent. They include three editors' invited articles (with discussion) in 2001: T.L. Lai (pp 303-408; Sequential analysis: some classical problems and new challenges); Cheng and Wu (pp 553-604; Factor screening and response surface exploration); Bickel and Kwon (pp863-960, Inference for semiparametric models: some questions and an answer). In addition, a special issue on bioinformatics, co-edited by Jun Liu (Harvard) and Bin Yu (Berkeley) appeared in Jan. 2002. A theme topic on statistical and mathematical finance, edited by Ruey Tsay (U of Chicago), is also about to appear in short time.

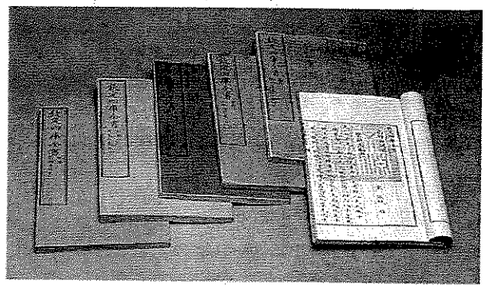
The acceptance rate for the regular submissions is about 25%. For over 70% of the submissions, the initial decisions were made within 6 months. Writing is still a major problem for many authors. A common misconception is the authors' unrealistic assumption that the reviewers should know the literature well enough to appreciate the importance of the topics addressed. Novelty of the approach and relevance to statistical practice are not highlighted; critical and routine steps are indistinguishable in theoretical derivations; scientific merits of the "real-life" examples are poorly discussed; lessons from simulation study are not clearly presented. Such reasons are common for rejection. Long papers are more difficult to review than short ones. They tend to lack focus, organization and conclusive results, thus appearing more like a set of research notes.

We have been extremely lucky to have Don Ylvisaker as the Consulting Editor. Don has spent so much time with each accepted paper. Don's editing is best described as a fantastic process of information filtering that helps bring out the statistical essence in the paper. In addition to standard grammar corrections and minor changes for the journal style, he did every thing to make sure that concepts and results are conveyed in the most elegant and precise manner, to such an extent that we often found a good portion of a section, which could be the introduction, the main derivation, the conclusion or even the technical appendix, was completely rewritten by Don.

Starting August 1, 2002, Hwai-Chung Ho (Academia Sinica, Taiwan) and Jane-Ling Wang (University of California at Davis) took over the editorial responsibility. A new website for electronic submission has been established, and they encourage authors to take advantage of this fast route. The new website for paper submission is: http://venus.stat.sinica.edu.tw/ss/author.htm.

Ker-Chau Li, Ph.D., is Professor of Statistics, University of California, Los Angeles, California, USA.
Email: kcli@stat.ucla.edu.

Yi-Ching Yao, Ph.D., is Professor of Statistics, Academia Sinica, Taiwan.
Email: yao@stat.sinica.edu.tw.



STATISTICA SINICA Vol. 12, No. 4, pp. 965-1290 October 2002

STATISTICA SINICA

Volume 12 Number 4 October 2002

Interview with a Distinguished Statistician

A Conversation with Dr. T. Y. Lee

By: Kao-Tai Tsai, Ph.D.

Dr. Lee holds a very special status among his fellow statisticians and within the pharmaceutical industries. Coming from a very humble root and through his hard work and perseverance, he successfully sailed his life across three challenging stages – academics, corporate statistician, and entrepreneur. Now he has stepped into another stage of his life – retirement. Yet he continues to be a mentor for young aspiring statisticians. It is indeed a privilege to have this opportunity to discuss with Dr. Lee various topics of work, life and humanity. To enjoy this story and important lessons of life, please read on...

Would you please tell us about your childhood and family background?

Before I reached the age of five, my family ran frequently from town to town in war-ridden China. With six children tagging along, my parents endured extreme hardships while escaping from Northern China, fleeing to the East Coast and finally across the sea to Taiwan in 1949. Life in Taiwan was harsh considering we had no possessions and my father's meager salary could barely support the eight of us. I still recall the struggles that my parents had attempting to borrow money to pay tuition for us. At one time I overheard my mother telling my father that she would not mind pawning her only wedding ring in order to cover our tuition. We had raised twenty chickens that produced about a dozen of eggs per day for us to sell in a local market. We woke up very early in the morning to buy rotten fish and cabbage peels at a local market. After mixing them with feed, we had to feed the chicken. We also gathered chicken droppings and sold them to the farmers as fertilizer before heading for school. After school, we helped my mother to make matchboxes so we could sell them. For every hundred matchboxes we produced, we made one dollar. We would usually make a couple hundred matchboxes before dinnertime.

During the college years, I held a part-time job maintaining the campus lawn to subsidize my living expenses. No modern lawnmower was available in early 1960s in Taiwan and the only tool I had was a small hand-held sickle. It took me whole afternoon to finish about half an acre of the lawn. It was a good test of both physical and mental endurance learned from my early childhood.

After I graduated from college, I received a scholarship and decided to pursue my graduate studies in the United States. Since I did not have money to come to America by air, I bought a slot in a cargo ship and arrived at Savanna, Georgia after about a month from port to port. After two days I arrived at Minneapolis by Greyhound bus in the cold winter March 1967. With a suitcase and a few books, I began my new life in the U.S.A.

What were your work experiences in Graduate School?

While I was a graduate student at University of Minnesota in 1967, I was very fortunate to be able to work as a research assistant in Mayo Clinic. The Clinic gave me the opportunity to gain experience in the fundamental activities of clinical trials, including CRF design, coding, keying, verifying, analyzing, report writing and consulting with the clinicians and epidemiologists. As a foreign student with limited language ability, I considered using English to interpret the meaning of the data for the clinicians a challenge. My grilling first year in Mayo clinic built a good foundation for the development of my career in the Pharmaceutical industry.

In 1970 I joined the PRO-CON project while studying at University of Pittsburgh under Dr. C.C. Li and Mr. Jerome Cornfield. It was an AID-sponsored (Agency for International Development) large-scale clinical trial testing the effectiveness of 1) prophylaxis against Gonorrhea, and 2) contraceptives. As the sole statistician on the team, I worked closely with professionals from various fields including chemistry, toxicology, pharmacology, clinical and epidemiology. This was the greatest training ground for learning skills that I've found to be truly valuable in the 'real world.' Through my PRO-CON experience, I grew in my ability to communicate clearly, work alongside very different personalities, and make independent decisions.

Please describe your statistical and managing experience in pharmaceutical industries.

My first-hand experiences in Mayo Clinic and in the PRO-CON project helped me to launch my pharmaceutical career starting with Merck in 1973. While I was a project statistician in 1976 at Hoechst-Rousell Pharmaceuticals, I was directing/managing a project consisting of more than two dozen studies of an analgesic and an anti-inflammatory compound. I was being challenged in the areas of managing multiple tasks, resource utilization, and time allocation. Later, I decided to pursue an Associate MBA program offered by the Wharton Business School of the University of Pennsylvania by taking night classes while working during the day. In 1982, I was named as Vice President of Biostatistics and Clinical Program of Ayerst Laboratories in Manhattan. I was responsible for statistical and data management supports for clinical research, pharmaceutical research and development, analytical research and development, quality control, toxicology, manufacturing, industrial hygiene, legal issues and pricing justification. In this position I was exposed to the numerous and varied elements of the entire drug development process, giving me a thorough understanding of the industry.

What motivated you to start a Contract Research Organization (CRO)?

In September 1987, I left Ayerst when it merged with Wyeth in Pennsylvania, and then started a CRO. At that time, the business world was going through rapid changes, dealing with issues such as competitiveness; number of usable years of a patent; unrealistic profit expectation from Wall Street; employment-at-will; higher frequency of merging, consolidation, and layoff; lack of mutual expectation of loyalty between company and employees; and the new concept of the "company without walls." The choice was clear. It was time to launch my own business! The American way would have been to start my company in the garage. Maybe because I was raised in Taiwan, I was free to break the tradition. So I started the company in my *attic* after taking a second mortgage on my house in 1988. Our company grew and expanded steadily and it was acquired by another CRO in 1998.

What have you learned in starting, managing, growing and leaving a business?

As I look back on my experience in the academics and industry, I recognized my weaknesses and my strength. Therefore, I like to share some of the experiences and things that I wish I had known that would have made the process of starting my own business a little easier... Here's what I've learned:

Preparation:

Before venturing into the entrepreneurial business, it is so critical to realize the *power of a good name*. A good professional name forces us to be better, and also helps the others to interact with us in the workplace or in the business negotiation. The best time to start establishing your reputation is when you are still young and employed. It is not too late to improve

yourself by being positive and honest, networking, improving social skill, breaking away from self-imposed limitations, team-building and learning the salesmanship. For the past 10 years, a new group of bright statisticians and programmers have joined the pharmaceutical companies. Many of them have a tendency to impose self-limitation to their career growth by not speaking English, working as a consultant instead of an employee to build relation and career in the work place. Their professional and work interactions are primarily limited to Chinese-speaking statisticians or programmers because of language and social barriers. I wish to encourage them to expand their networking to all levels of professional and racial groups. "We don't just make lemonade from a few lemons; we need to grow a grove of lemon trees."

In addition to equip yourselves at the workplace and at the community level, you need to find out whether your spouse has the emotional and inner strength to go through the challenges in starting a business.

It is imperative to know your niche. Are the key players of your conceptual company well known in the industry and professional society? Do you have a good trade name and network to explore the possibility of project award in relatively short time after launching the business? Design your marketing brochure to highlight the niche and the strength of your company. Know how to price your services according to the quality of your services and timeline. To start with a sound financial plan, it is important to have the professional advises from accounting, legal, financial planning and banking professionals. Prepare a good and realistic 5-year business plan. You need to have the cash to last you for at least one year and be prepared to have low or no salary for yourself.

1. The time to launch:

Opportunity does not gather dust, therefore, be decisive to take the first step. You have to be willing to make sacrifice and eat and sleep with your company. Dual pressures of marketing in daytime and contract fulfillment at night will generate the feeling of loneliness and self-doubt. This is the reason why a supportive spouse, your inner strength and faith are so critical to sustain your business.

2. Sustain and grow the business:

As a private company grows in size, there are several important factors that need to be taken into consideration. These factors are: the vision and the mission statement of the founder and the company, business environment, the cost of strengthening the infrastructures to manage the growth, the cost of financing and maintaining a public company, the constant pressure of fiduciary responsibilities to the share-holders, the change in life style and the responsibilities toward our own family, church and the community. Success is achieved in inches, not miles and only 7% of all start-up companies will survive beyond 3 years. As the company grows, we have to let the system do the job of supervising and quality control. The founder and the key players will have to set a good example in character and integrity and to foster trust among employees. The most difficult task is to keep everyone's ego in check and balance the creative conflicts and maintain internal harmony. To deal with unreasonable clients, the founder or the key players have to make themselves available to resolve disagreement. It is a lot easier just to tell the truth when you have screwed up and make proper adjustment to the client. Radiate your passion and enthusiasm to your colleagues and clients, and be a director not a dictator. Find unusual ways to build team spirit and rid off the routine blue. Give frequent and appropriate awards to employees to offer encouragement and incentives.

As business expands, internal expectation of benefits, perks and salary grows even faster. Pricing strategy should keep up with the business expenses.

3. Strategic direction of a corporation:

The management has to make tough decision with respect to the strategic direction of the corporation. Basically there are three choices: maintain status quo and grow, go public or merge/ acquire.

4. Preparation for merge/IPO:

a. Engaging the business-consulting firm to make steps for building up the name recognition of the company. Make sure that the infrastructure and the succession plan are in place. The founder and the key players should prepare the clients and assure them that their projects would not suffer during or after the merge/IPO.
b. Legal identity of the corporation: It involves the selection of the type of the legal incorporation. The management needs to be aware of the impact of changing from the S-corporation to C-corporation. The company book should have detailed and accurate records of the board decisions and signed by the board members. The financial report should be prepared by reputable accounting firm and audited by an independent firm.
c. Outside financial consulting group should be hired to determine the fair market value of the corporation and a Descriptive Memorandum (DM) of the corporation has to be prepared. The consulting firm will distribute the DM to the potential under-writers or partners.
d. At the same time, the management can not afford to ignore the clients and has to concentrate in running the business. Therefore, a team of lawyers, financial consultants, accountants and communication specialist should be hired to undertake the lengthy process of IPO or merger.
e. Careful and lengthy process will be required to conduct due diligence, to discuss issues face to face, to produce confidentiality agreement and to communicate the progress to employees and clients.
f. Infusion of cash or stocks or the combination of both to the company should be planned carefully with the financial consultants. You need to understand the pitfalls of 338(h) (10) and a good lawyer should help to negotiate the stock restrictions. Since the public company will be under constant scrutiny, it is important to negotiate in advance, reasonable employment letters for all employees, exit options and to define the liability.

5. After the dust settled (for me):

Time and tide wait for no man. It is not the end of my active participation in the society after retirement. On the contrary, I have forged my energy and financial resources to support selected non-profit organizations and Christian mission organizations, to do volunteer work and to be available as a mentor to young people. Gardening has become one of the most relaxing and rewarding activities in my summer time. Though outwardly I seem to be wasting away after retirement, yet inwardly I am being renewed day by day because of the Christian faith I have. A new chapter of my life has just begun.

Do you have any advises for the young statisticians?

My experience has been very limited and unique to my own background and faith. I can only share what I have learned:

- Pay attention to social etiquette and learn to work well with colleagues in the office. You want to build a good trade name, which is not established in a day.

Interview with Dr. T. Y. Lee ...

- The pharmaceutical industry is "a small world." Expand networking to all levels of professionals because you need friends (particularly if you intend to start your own company in the future.)
- Strive to be positive, ethical, and honest and always offer a helping hand to others.
- Take time to listen.
- Be a team player and share your sense of mission toward company goal.
- Define your own mission statement. Know what you want in life and do the best you can in all you do.
- Your family should always be at top of your priorities. Build a good marriage and teach your children integrity, love, faith and hope.

Before we close this conversation, I would like to share a quotation from Mr. Art Linkletter. This quotation has a very good description of how I believe we should live out our day-to-day life:

"Do a little more than you are paid to,
Give a little more than you have to,
Try a little harder than you want to,
Aim a little higher than you think possible,
And give a lot of thanks to God for health,
family and friends."

❑

Painting at right: A Scholar's Retreat
By Master Artist: Chang-Shi Wu

The majority of classical Chinese paintings exhibited landscape such as mountains, rivers, and pine trees, which signified kindness, wisdom, and forever prosperity. In this painting, friends enjoy their time together by playing chess, drinking, and composing poetry.

Special Feature Article:
Monte Carlo & Bioinformatics

By: Jun Liu, Ph.D.

Dr. Liu, Professor of Statistics at Harvard University, is the recipient of the COPSS Presidents Award in 2002. We are extremely honored to have him share with us about his research, his interests, and his accomplishments. In this *Special Feature Article*, we have the privilege to publish his acceptance speech of the COPSS award and his adventure in the field of Monte Carlo and Bioinformatics. – *The Editor*

The Acceptance Speech of The COPSS Award, 2002 - Jun Liu

Thank you, Steve, for that tremendous introduction, thank you, the committee, for this great honor, and thank you all for being here to share this moment with me.

In those old graduate-school days, we used to sit at the back row of the classroom, admiring the names on the COPSS Award list. I never imaged that one day I could be standing here, receiving this great honor, and having my name on that list; but we did figure out that "Peter" is perhaps a good name to have.

This has been a tremendous year for me. First, my son James was born on February 9th, at 10:22PM, that is, 22:22. Then, in March I was told the winning of the COPSS Award, the 22nd award! Interesting coincidence.

I know I am a very lucky guy, but I did not realize that I am SO much blessed.

I am very lucky to have chosen statistics as my life-long career, and I am grateful to the people who have helped (or forced) me with that decision. Statistics has allowed me the opportunities to be both indulged in mathematics and fascinated by science. The longer I stay in this field, the more I love its people and the more I admire its upbeat yet benign spirit. Unlike in some other fields, we can more or less freely exchange ideas, read others' technical reports and let others read ours, from which I have benefited a lot; unlike in some other fields, we can be very versatile and work on many different subjects and are still under the big "statistics" umbrella, from which I have enjoyed a lot. Because of these collaborative nature and benign spirits, much of my work results from intensive collaborations with other people. Thus, this great honor ought to be shared by all my collaborators, especially Rong Chen, Augustine Kong, Chip Lawrence, Wing Wong, and many of my students.

No lawn looks nice without being cared and mowed, and no life grows great without being loved and disciplined. I am very lucky to have met many great people who have influenced me in one way or another; they collectively shaped my life and defined who I am. I thank all those people who have loved me, encouraged me, and helped me; all those people who have taught me, entertained me, and inspired me; all those people who have harnessed me, disciplined me, and perhaps even made decisions for me.

Other than my parents, the person who influenced me the most while I was pondering over my future at age 22 was Professor Bill Strawderman. He not only taught me a beautiful course on mathematical statistics but also wrote me a fabulous recommendation letter enabling me to go somewhere else --- such as the

Special Feature Article ...

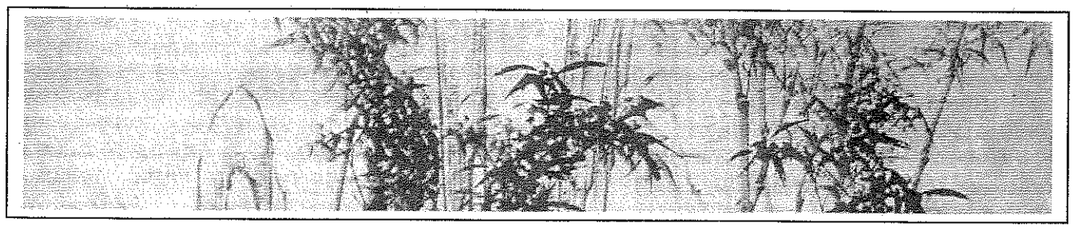
University of Chicago. In Chicago, I had the honor and luck to work closely then, and ever since, with my thesis advisors Professors Wing Wong and Augustine Kong. They taught me many things including how to think critically and scientifically, how to formulate a meaningful research question, and how to put one's heart in what he does. Their devotions to science are always my inspiration. Professor Don Rubin at Harvard was both my first boss and my post-PhD mentor and my current boss. He not only taught me how to think about statistics beyond its mathematical components, but also hired me repeatedly. Professor Persi Diaconis has been always inspiring, supportive, and nurturing to me since I first met him at the University of Chicago. Besides teaching me about group theory and Markov chains, he persuaded me to spend 6 wonderful years being nurtured in another great Statistics Department ---- located in the famous west campus of Harvard. During my 6-years' stay, my Stanford colleagues have set me a very high standard through their own doings. I learnt from them that dedicated statisticians should always be inspired to challenge themselves. Dr. Chip Lawrence from the Wadsworth lab is responsible for leading me into the wonderful new land of bioinformatics. I have enjoyed immensely the fruitful collaborations with him in the past 10 years.

I am very lucky to have a pair of wonderful parents. They taught me to love science, to be honest, and to be focused. I remember that they spent days and nights transcribing a thick mathematics book for biked over the whole city trying to buy some science books for me when I was 13. I also remember that they made me come to the US to seek after a higher goal when I was 21. Without their love, their directions, and their true belief in me, I would have gone nowhere.

I am very lucky to have a wonderful wife Wei. I am most grateful to her for her love, her inspiration, her wholehearted support, her patience, and her persistence in the past nine years. She has made me a more complete and caring human being; and a stronger, better, and cleaner man. My son James is a recent joy of my life. I'll report to you later how he becomes a source of my inspirations.

Yesterday, August 13th, was the 16th anniversary of my first arrival in the US. Sixteen years ago exactly today, I walked on the street of New York City for the first time in my life. While I walked on the 42nd, I felt like being in a movie. That was my first time to see so closely so many tall buildings lumping together --- that was my first impression of the city, and of the US. Sixteen years later, in the very same city and on the very same day, I am receiving one of the highest honors in the statistical science profession. The United States of America is perhaps the only country in the world that can make this possible. I love American spirit and I love our statistics profession. Thank you.

Jun Liu, Ph.D., is a Professor of the Statistics Department, Harvard University, Cambridge, Massachusetts, USA.
Email: jliu@bioinfo.stat.harcard.edu



Why I am interested in Monte Carlo and Bioinformatics

Jun S. Liu
Department of Statistics
Harvard University

For many years since childhood, I have been fascinated by mathematical problems. But it is my parents who allowed me to indulge myself in the wonderland of mathematics, fostered my confidence in solving difficult problems, and kindled my love of natural sciences. I received the B.S. degree in Mathematics in 1985 from Peking University in Beijing. In 1986, I was admitted to the Ph.D. program of mathematics at Rutgers University. After staying in the program for a year or so, I was struck and tortured by one fundamental question: why should I spend many years of my life solving that problem my advisor gave me? So I decided to pursue something that is more relevant to real life and easier for me to judge its value. By chance I chose statistics as my life-long career, and by luck I was admitted to the University of Chicago and studied with Professor Wing Hung Wong. After getting my Ph.D. degree in 1991, I took an Assistant Professor position in the Statistics Department of Harvard. During my stay at Harvard, I learned a lot from the great statisticians Don Rubin, Art Dempster, and Carl Morris. Being attracted by its great fame and its closeness to China, I accepted a tenure-track Assistant Professor offer from the Statistics Department of Stanford University in 1994. In 2000, I returned to Harvard Statistics Department as a full Professor. My research interests focus on two major areas interspersed with smaller but related topics: Monte Carlo methodologies and computational biology/bioinformatics. I will explain why they have been so fascinating to me.

1 Monte Carlo

Monte Carlo is one of the most versatile and fascinating topics in mathematical sciences. An essential part of many scientific problems is the computation of integral $I = \int_D g(\boldsymbol{x})d\boldsymbol{x}$, where D is often a region in a high-dimensional space and $g(\boldsymbol{x})$ is the target function of interest. If we can draw iid random samples $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(m)}$ uniformly from D (by a computer), an approximation

to I can be obtained as

$$\hat{I}_m = \frac{1}{m}\{g(\boldsymbol{x}^{(1)}) + \cdots + g(\boldsymbol{x}^{(m)})\}.$$

The *law of large numbers* states that $\lim_{m\to\infty} \hat{I}_m = I$ with probability 1 and the *central limit theorem* gives its rate of convergence: $\sqrt{m}(\hat{I}_m - I) \Rightarrow N(0, \sigma^2)$, where $\sigma^2 = \mathrm{var}\{g(\boldsymbol{x})\}$. This basic setting underlies the potential role of the Monte Carlo methodology in science and statistics.

Although the "error rate" of the Monte Carlo approximation is $O(m^{-1/2})$, regardless of the dimensionality of $\boldsymbol{x}$, two intrinsic difficulties arise: (a) when the region D is large in high-dimensional space, the variance σ^2, which measures how "uniform" the function g is in region D, can be formidably large; (b) one may not be able to produce uniform random samples in an arbitrary region D. To overcome these difficulties, researchers often employ the idea of *importance sampling* in which one generates random samples $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(m)}$ from a nonuniform distribution $\pi(\boldsymbol{x})$ that puts more probability mass on the "importance" region of the state space D. One can then estimate integral I as

$$\tilde{I}_m = \frac{1}{m}\sum_{j=1}^{m} \frac{g(\boldsymbol{x}^{(j)})}{\pi(\boldsymbol{x}^{(j)})},$$

which has a variance $\sigma_\pi^2 = \mathrm{var}_\pi\{g(\boldsymbol{x})/\pi(\boldsymbol{x})\}$. In the most fortunate case, we may choose $\pi(\boldsymbol{x}) \propto g(\boldsymbol{x})$ when g is non-negative and I is finite, which results in an exact estimate of I. But in no known real problem has this "luckiest situation" ever occurred. More realistically, we try to find a good "candidate" π which will explore more in the "importance" region of g. In such a situation, generating random draws from π can be a challenging problem.

I often tell my students that if they can invent a method to draw good random samples from an arbitrary probability distribution, they will be able to get a number of Nobel prizes, in addition to starting up a few companies with the potential of Intel and Microsoft. The simple reason for this absurd statement is that many fundamental problems in physics, chemistry, and biology, and many combinatorial optimization problems can be formulated as high-dimensional simulation problems, for which the Monte Carlo method is one of the most powerful tools. As summarized in my recent book (Liu 2001), my own research in Monte Carlo methods hinges on two general strategies and their applications: Markov chain Monte Carlo (MCMC) and sequential Monte Carlo (SMC). My focuses are on theories that enhance our understanding of the basic principles underlying a good Monte Carlo algorithm and ideas that enable us to design more efficient ones.

1.1 Markov chain Monte Carlo

MCMC is by far the most effective means for sampling from a high-dimensional distribution. Its basic idea to to simulate on a computer a Markov chain whose stationary distribution is the target one π. The Metropolis algorithm and the Gibbs sampler are two general-purpose recipes for designing the transition rules for the chain, but other more specialized and exquisite transition rules have also been proposed (see Liu 2001). These transition rules are critically important for the efficiency of the resulting algorithm. I have been mostly interested in the following general issues: (a) the design of more efficient MCMC algorithms; (b) convergence bounds and convergence diagnostics; (c) the comparison of different MCMC schemes; (d) efficient use of Monte Carlo samples; and (e) novel applications of MCMC techniques.

In my thesis, I studied several related theoretical properties of the Gibbs sampler and derived a novel representation of the autocovariance function. This representation was used to compare different algorithms and establish the geometric convergence property of the Gibbs sampler (Liu, Wong & Kong 1994, 1995). The representation is also suggestive for designing more efficient samplers (e.g., grouping and collapsing). I later generalized these results and proposed a Gibbs sampling algorithm for biological sequence alignment (Liu 1994), which becomes quite popular in recent years among bioinformatics researchers. I have explored ways of generalizing the Gibbs sampler — our transformation-group viewpoint is particularly attrative (Liu & Wu 1999; Liu & Sabatti 2000). This new theory helps one understand and unify several recently proposed improvement strategies (e.g., multigrid Monte Carlo and parameter expansion) for MCMC sampling and enables one to conduct efficient conditional sampling in a very general sense. A similar exploration of the Metropolis algorithm led to a new Metropolis-type transition rule, which can help a MCMC chain jump out of local mode more easily , and a theoretical underpinning of the dynamic weighting method (Liu, Liang & Wong 2000, 2001).

1.2 Sequential Monte Carlo for dynamic systems

The sequential Monte Carlo (SMC) methodology recently emerged in the fields of statistics and engineering has shown a great promise in solving a large class of highly complex inference and optimization problems, opening up new frontiers for cross-fertilization between statistical science and many application areas. SMC can be loosely defined as a family of recursive Monte Carlo techniques for solving on-line estimation and prediction problems in stochastic dynamic systems. By recursively generating Monte Carlo representations (i.e., random draws) of the state

variables, SMC adapts flexibly to the dynamics of the underlying stochastic system, regardless whether it is linear and Gaussian or not. These techniques have found wide applications in solving many scientific and engineering problems (Chen et al. 2000, Doucet et al. 2001, Liu 2001).

In many problems, it is of interest to evaluate the expectation of a function $h(\boldsymbol{x}_t)$ with respect to π_t, or to find the mode of π_t. One of the most commonly seen formulation is the *generalized state space model* of the form

$$\begin{array}{rlll} \text{(state equation):} & x_t = \phi_t(\boldsymbol{x}_{t-1}, \varepsilon_t) & \text{or} & x_t \sim f_t(\cdot \mid \boldsymbol{x}_{t-1}) \\ \text{(observation equation):} & y_t = \gamma_t(\boldsymbol{x}_t, e_t) & \text{or} & y_t \sim g_t(\cdot \mid \boldsymbol{x}_t), \end{array} \tag{1}$$

where $\boldsymbol{x}_t$=$(x_1, \ldots, x_t)$ is the unobserved state variable, $\boldsymbol{y}_t$=$(y_1, \ldots, y_t)$ is the observed data up to time t, ε_t and e_t are the noises that lead to the conditional densities f_t and g_t. It includes as special cases the *hidden Markov Model* (HMM) widely used in speech recognition and bioinformatics, popular time series models in engineering and financial data analysis; and the dynamic Bayesian network in the AI community. The optimal on-line estimation (in terms of the mean-squared error) of $h(\boldsymbol{x}_t)$ is $E[h(\boldsymbol{x}_t)|\boldsymbol{y}_t]$, where the expectation is taken with respect to the time-t posterior distribution $\pi_t(\boldsymbol{x}_t) = p(\boldsymbol{x}_t|\boldsymbol{y}_t)$. For linear and Gaussian systems, $\pi_t(\cdot)$ is Gaussian and computed recursively by the Kalman filter. When x_t only takes on a few discrete values, a forward-backward method can be used. In all other cases, this Bayes solution cannot be computed exactly and SMC has been shown an effective means to combat the complexity.

A central component of the SMC is sequential importance sampling (SIS). For example, although the goal in the above example is to sample $\boldsymbol{x}_n$ from π_n, we may proceed by sampling x_1 from a trial distribution $q_1(x_1)$, x_2 from $q_2(x_2|x_1)$, ..., and $q_n(x_n|\boldsymbol{x}_{n-1})$ and then weight the draw by

$$w(\boldsymbol{x}_n) = \frac{\pi_n(\boldsymbol{x}_n)}{q_1(x_1)q_2(x_2|x_1)\cdots q_n(x_n|\boldsymbol{x}_{n-1})}.$$

This is exactly the importance sampling except that the sampling distribution is built up sequentially. Since in most applications we can evaluate $\pi_n(\boldsymbol{x}_n) = p(\boldsymbol{x}_n \mid \boldsymbol{y}_n) \propto p(\boldsymbol{x}_n, \boldsymbol{y}_n)$ only up to a proportional constant, the draw $\boldsymbol{x}_n$ is weighted in reality in proportional to

$$w'(\boldsymbol{x}_n)\frac{p(\boldsymbol{x}_n, \boldsymbol{y}_n)}{q_1(x_1)q_2(x_2|x_1)\cdots q_n(x_n|\boldsymbol{x}_{n-1})}.$$

Note that $E[w'(\boldsymbol{x}_n)] = p(\boldsymbol{y}_n)$, implying that the average of these weights is an unbiased estimate the Bayes factor.

In the sequential imputation method, Kong et al. (1994) chose $q_t(x_t|\boldsymbol{x}_{t-1}) = \pi_t(x_t|\boldsymbol{x}_{t-1}) = p(x_t \mid \boldsymbol{x}_{t-1}, \boldsymbol{y}_t)$, and in the bootstrap filter, Gordon et al. (1993) chose $q_t(x_t \mid \boldsymbol{x}_{t-1}) = p(x_t \mid \boldsymbol{x}_{t-1})$. It was soon realized (Liu & Chen 1995) that carrying out the above SIS scheme all the way through is not ideal in many applications. Typically, many sampled $\boldsymbol{x}_n$ land in unimportant regions of π_n, resulting extremely small weights. A key development to rectify the problem is to conduct occasional resampling in a parallel implementation of the SIS. More precisely, for $t = 1, \ldots, n$, a parallel SIS completes the following two steps:

(A) For each j, $j = 1, \ldots, m$, generate a $x_{t+1}^{(j)}$ (or multiple of them) from $q_{t+1}(x_{t+1} \mid \boldsymbol{x}_t^{(j)})$; attach it to $\boldsymbol{x}_t^{(j)}$ to form $\boldsymbol{x}_{t+1}^{(j)} = (\boldsymbol{x}_t^{(j)}, x_{t+1}^{(j)})$. Each $\boldsymbol{x}_t^{(j)}$ is called a *"stream."*

(B) Compute the "incremental weight"

$$u_{t+1}^{(j)} = \frac{\pi_{t+1}(\boldsymbol{x}_{t+1}^{(j)})}{\pi_t(\boldsymbol{x}_t^{(j)})q_{t+1}(x_{t+1}^{(j)} \mid \boldsymbol{x}_t^{(j)})}; \quad \text{and let} \quad w_{t+1}^{(j)} = u_{t+1}^{(j)} w_t^{(j)}. \tag{2}$$

A resampling step, inserted between recursions steps (A) and (B) periodically, draws a new sample $\{\boldsymbol{x}_t^{(*j)}\}_{j=1}^{m^*}$ from the current one $\{\boldsymbol{x}_t^{(j)}\}_{j=1}^{m}$ according to a probability vector $\{a_t^{(j)}\}_{j=1}^{m}$. Each new draw is assigned a new weight $w_t^{(j)}/a_t^{(j)}$ so that the sample remains properly weighted. Resampling provides opportunities for the good streams to replicate themselves, steering the algorithm towards promising directions (Liu & Chen 1998). A graphical representation of the SMC method is shown in Figure 1.

In order to design an efficient SMC scheme (in terms of both accuracy and low complexity), one needs to consider many important issues. For example, it is important to configure an SDS properly (e.g., choosing a good parameterization) so as to result in an efficient SMC algorithm. The key is to balance accuracy and complexity/difficulty. Choosing good trial distributions q_t is one of the most critical steps. A basic principle is to use as much available information (observation) in designing q_t as the computation allows. It is also useful to anticipate future development of the system and design q_t accordingly. We have also shown that marginalization (integrating out some components) can improve an SMC's efficiency (MacEachern et al. 1999). Based on this principle, we developed a mixture Kalman filter (MKF) for conditional dynamic linear systems (CDLM) (Chen & Liu 2000), in which can greatly increase the efficiency of SMC. Another difficult problem encountered in SMC is the impoverishment of distinct streams/particles (Doucet et al. 2001). Several ideas for generating new streams along with SMC recursions have been proposed and shown effective in practice. These include MCMC-based moves, local perturbations,

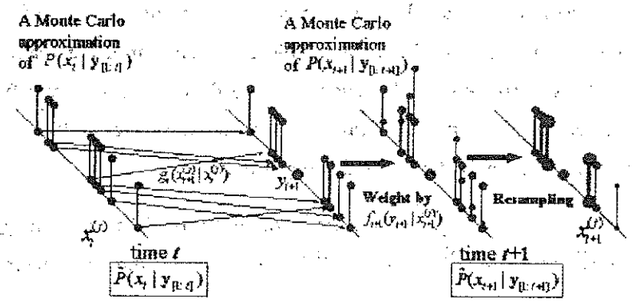


Figure 1: Graphical Illustration of Sequential Monte Carlo. Suppose all the Monte Carlo samples have equal weights at time t. At time t+1, the new samples are generated by the state equation, weighted by their predictability of y_t, and then resampled.

rejection control, etc. (Doucet et al. 2001, Liu 2001). Although the implementation of these ideas often involves in-depth understanding of the specific problems of interest, the development of general guiding principles is the key to the recent surge of interest in SMC technologies.

The applications of SMC are almost as versatile as MCMC (Liu 2001). For example, in studying chain polymers, the scientist often employs the self-avoid walk (SAW) model. It has been shown (Zhang & Liu 2002) that SMC can be applied effectively to find the minimal energy configuration of such a model by adding one (or a few) monomer a time. As another example, In some applications such as ecology (Sanderson, 2000), one is interested in counting or simulating uniformly the $m \times n$ tables that contain only 0's and 1's and have the fixed row and column marginal sums. A very simple SMC method was developed by Y Chen in his Ph.D. thesis (2001, Stanford), in which we showed that the efficiency gain over the MCMC competitors can be many order of magnitudes (Liu 2001). Briefly, this approach starts with the first column of the table and then proceed through all columns sequentially. In each step, it samples the positions of 1's in that column with a certain probability vector and then update the row sums accordingly. The application of SMC in bioinformatics such as sequence analysis is also being explored.

2 Computational biology

A wealth of biological sequence data and microarray expression data has emerged from the human genome project and functional genomics studies and has fostered the bioinformatics hype among the statisticians. *In Silico* methods for understanding these data and for incorporating different sources of biological information/knowledge are becoming increasingly important. Because the nature is inherently stochastic, the essence of much of the computational effort is statistical data analysis and probabilistic modeling.

The word "genome" refers to the collection of all the chromosomes (i.e., chains of DNA bases; human has 23 pairs of these) in a cell. Certain segments of the genome called "genes" (or coding regions) encode the information needed to make proteins, which are action molecules of the cell, responsible for nearly all cellular processes. It is estimated that the human genome has about 30,000 genes accounting for ~3% of the genome. The expression of these genes, i.e., the amount of protein products to be made in a cell, is tightly regulated so as to meet the requirements of specific cells and for cells to respond to changes in their environment. A central goal of molecular biology is to understand how protein synthesis is regulated (Lodish et al. 2000). The main focus of my research group is on the design of effective statistical models and computational strategies for understanding biological and genetic data. We are particularly interested in the following topics: (a) predicting gene regulatory binding motifs; (b) homology modeling and sequence-based protein analysis; and (c) human SNPs and linkage disequilibrium studies. I will discuss some of these problems below.

2.1 Microarrays and Discovery of Cis-Regulatory Binding Motifs

In order to make a protein molecule, a gene is first transcribed to messenger RNA (mRNA), an easily degradable molecule, which then carries the information to a cellular machinery (ribosome) for protein production. While there are several levels of gene regulation, the dominant form is transcriptional regulation. Specific sequence signals upstream of each gene provide a target, called the promoters, for RNA polymerase (a machinery for transcription) to bind so as to initiate the transcription. Usually, however, some proteins called *transcription factors* (TFs), which specialize in gene regulation, will first bind to their target sites (called the TF binding sites) near the promoter of a gene and interfere with the RNA polymerase, either repressing or enhancing the production of mRNA. A TF can regulate many genes and its binding sites (the

length w of these sites range from 7 to about 20 bases) often exhibit a conserved pattern, just like that the TF recognizes its preferred word.

A interesting and challenging problem is the computational identification of the TF binding sites, and their implied pattern (termed as a "binding motif"). From clustering analysis of microarray experiments, one can first gather sets of genes that are likely regulated by common TFs (Roth et al. 1998). Then, one can search the upstream regions of the genes (the few hundreds bases before the gene start) in each cluster for significant recurring patterns. A pattern can be an exact "word" or a fuzzy word represented by a $4 \times w$ *weight matrix* Θ, where its jth column represents the frequencies of the four bases occurring at the jth position of the motif. Lawrence & Reilly (1990) first introduced a formal statistical model and designed an expectation-maximization (EM) algorithm for the discovery of TF binding sites, with the assumption that every gene in the cluster must contain one TF sites in its upstream region. Bailey & Elkan (1994) later extended this EM method to Multiple EM for Motif Elicitation (MEME), which relaxes the one-site-per-sequence condition.

With collaborators, I have developed several Gibbs sampling-based motif finding methods including the site sampler, the Gibbs motif sampler (GMS), BioProspector, and MDscan (Lawrence et al. 1993, Liu et al. 1995, Neuwald et al. 1995, Liu et al. 2001, Liu et al. 2002). The GMS starts with a collection of non-overlapping words of length w (called w-mers) randomly selected from the sequence dataset. The total number of the w-mers is determined by a prior motif abundance distribution provided by the user and is updated throughout the iteration. The initial weight matrix is built based on the alignment of these w-mers. At each iteration, the algorithm uses the current weight matrix to score every position in all the sequences, deciding stochastically whether it should be regarded as the start of a TF site. The weight matrix is updated once a new site is sampled in or an old site is kicked out. Other Monte Carlo moves such as fragmentation and phase-shift have also been designed to enable the algorithm to better explore the likelihood surface (Liu 1994). There are two distinct advantages of the GMS: (a) it is based on a Bayesian formulation and the underlying model can be further enriched to accommodate more complex transcription regulation problems in high eukaryotes; (b) its Monte Carlo strategy is very flexible and can be further improved by incorporating the recent advances in Monte Carlo optimization techniques.

By adopting the GMS approach, Roth et al. 1998 (AlignACE) successfully discovered many experimentally determined TF motifs from yeast gene expression clusters. In a study of a set of

muscle-specific genes, Wasserman et al. (2001) first applied a Bayesian alignment method (Zhu et al. 1998) to discover highly conserved parts between Human and mouse intergenic regions and then correctly identified, using GMS, the binding motifs of three major muscle-specific TFs. McCue et al. (2001) combined the GMS with a cross-species comparison strategy to predict 2000 TF binding sites in E. Coli with 80% accuracy.

We recently developed two new motif discovery methods based on the Bayesian formulation. Compared with earlier GMS and AlignACE, the new methods achieve an even better sensitivity, flexibility, and speed. BioProspector (Liu et al. 2001) was able to model correlated sequence background and to accommodate the simultaneous occurrences of two co-linear motifs with flexible gaps (such as the TATA box). MDscan was designed to handle data produced by the Chromatin-Immunoprecipitation microarray experiments (Liu et al. 2002). It combines the advantages of an exhaustive word-enumeration strategy and the Bayesian formulation in GMS. It achieves a comparable sensitivity as BioProspector but is many times faster. We are now developing strategies to efficiently combine the information from the sequence data, cross-species comparisons, and microarray analysis; and to design new statistical models for eukaryotic gene regulation modules.

2.2 A broader array of bioinformatics problems

Microarray analysis provides for statisticians an excellent entry point to bioinformatics/ computational biology. Here I give a brief personal account on other bioinformatics challenges that await statisticians' contributions.

The protein folding problem, i.e., the prediction of the three-dimensional fold of a protein molecule based only on its primary sequence information, is often regarded as the crown jewel of the biopolymer research. Knowledge on the structures of target proteins and on how they interact with ligands is of great importance to drug designers. Although the 3-D structures of many proteins have been worked out by X-ray crystallographers, these structures only account for a small part of the protein universe and scientists are still not capable of predicting protein tertiary structures *ab initio*. Recently, theoreticians have turned their attentions to much simpler black-white bead model for understanding the design principles of protein structures (Zhang and Liu 2002). Practitioners have opted to use more statistically based threading method (Xu et al. 2002).

Multiple sequence alignment (MSA) is still the main tool for protein sequence analysis, which

has been at the center of computational biology for about 30 years. By extracting common conserved features from multiple related sequences, MSA helps one detect remote relationships among proteins and identify sequence positions of funtional and structural importance. With the completion of the human genome and genomes of many other species, the task of organizing and understanding the generated sequence and structural data becomes even more pressing and challenging. Many statistical and computational methods for sequence alignment has been proposed over the years, among which the most popular ones include Clustal W. (Thompson et al. 1994), PSI-BLAST (Altschul et al. 1997), SAM (http://www.cse.ucsc.edu/research/compbio/sam.html), and HMMER (http://hmmer.wustl.edu), etc. In particular, the application of hidden Markov models (Krogh et al. 1994) and the Gibbs sampler (Lawrence et al. 1993, Neuwald et al. 1995, 1997) to biopolymer sequence analysis has revolutionized the field. Pfam database (Bateman et al. 2002) contains a large collection of annotated protein family profiles built based on hidden Markov models and is becoming very influential in protein research. An emerging challenge is the analysis of aligned protein sequences in order to gain further insights on protein functions (Neuwald et al. 2002).

There have been some recent interests in incorporating gene ontology (GO) in microarray analyses. Gene ontology refers to a dynamically controlled vocabulary that can be applied to (the genes of) all organisms. Each gene product can be described by its molecular function (e.g., transcription factor), its involvement in biological processes (e.g., mitosis), and its cellular location (e.g., nucleus). Bringing GO into the analysis of high throughput biological data such as microarrays can be extremely insightful. Recently, in the analysis of circadian gene regulation, Storch et al. (2002) mapped various clusters of genes based on their microarray experiments to GO hierarchies and found that clock-regulated genes in heart and liver participate in many related processes even though the two sets of genes have almost no overlap.

3 Final Remarks

Physicist Freeman Dyson has called the 21st century the century of biology. I think that this is also a century of statistics. Besides the biotechnology revolution, exciting developments in other areas of science and technology have also generated huge amounts of data, e.g., single-molecule experiments in chemistry, galaxy images in astronomy, internet traffic monitoring, environmental modeling, etc. Statistics has never been so much sought after by biologists,

medical researchers, computer scientists, and many other scientists. But it is also a challenging century for all statisticians. Indeed, if we statisticians do not proactively participate in the information technology revolution, other scientists (e.g., computer scientists) will learn and do statistics whether we approve it or not. We clearly have an advantage for now, and we still can control our fate if we try.

References

S. F. Altschul, T. L. Madden, A. A. Schaffer, J. H. Zhang, Z. Zhang, W. Miller, and D. J. Lipman (1997). Gapped blast and psi-blast: a new generation of protein database search programs. *Nucleic Acids Research*, 25(17):3389–3402.

T.L. Bailey and C. Elkan (1994). Fitting a mixture model by expectation-maximization to discover motifs in biopolymers. *Proc. Int'l Conf. Intell. Syst. Mol. Biol.*, 2: 28-36.

Bateman A, Birney E, Cerruti L, Durbin R, Etwiller L, Eddy SR, Griffiths-Jones S, Howe KL, Marshall M, Sonnhammer EL. (2002). The Pfam protein families database. *Nucleic Acids Res*, 30: 276-80.

Chen, R. and Liu, J.S. (2000). Mixture Kalman filters. *J. Roy. Statist. Soc.* B, **62**, 493-508.

Chen, R., Wang, X., and Liu, J.S. (2000). Adaptive joint detection and decoding in flat-fading channels via mixture Kalman filtering. *IEEE Trans. Info. Theory*, **46**, 2079-2094.

A. Doucet, J. F. G. de Freitas, and N. J. Gordon (2001). *Sequential Monte Carlo in Practice.* Springer-Verlag, New York.

Gordon, N.J., Salmon, D.J. and Smith, A.F.M. (1993). A novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proc. Radar Signal Processing*, **140**, 107-113.

Kong, A., Liu, J.S. and Wong, W.H. (1994). Sequential imputation method and missing data problems. *J. Amer. Statist. Assoc.*, **89**, 278-288.

A. Krogh, M. Brown, I.S. Mian, K. Sjolander, and D. Haussler (1994). Hidden markov models in computational biology: Applications to protein modeling. *J. Mol. Biol.*, 235:1501-31.

Lawrence, C.E., Altschul, S.F., Boguski, M.S., Liu, J.S., Neuwald, A.F., and Wootton, J. (1993). Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment. *Science*, **262**, 208-214.

C.E. Lawrence and A.A. Reilly (1990). An expectation-maximization (EM) algorithm for the idenification and characterization of common sites in unaligned biopolymer sequences. *Proteins*, 7: 41-51.

Liu, J.S. (1994). The collapsed Gibbs sampler with applications to a gene regulation problem, *J. Amer. Statist. Assoc.*, **89**, 958-966.

— (2001). *Monte Carlo Strategies in Scientific Computing*. Springer: New York.

Liu, X., Brutlag, D.L., and Liu, J.S. (2001). BioProspector: Discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes. *Proceedings of the Pacific Symposium on Bioinformatics*, **6**, 127-138.

— (2002). An algorithm for finding protein-DNA binding sites with applications to chromatin immunoprecipitation microarry experiments. *Nat. Biotech.*, **20**, 835-39.

Liu, J.S. and Chen, R. (1995). Blind deconvolution via sequential imputations. *J. Amer. Statist. Assoc.*, **90**, 567-576.

— (1998). Sequential Monte Carlo methods for dynamic systems. *J. Amer. Statist. Assoc.*, **93**, 1032-1044.

Liu, J.S., Liang, F., and Wong, W.H. (2000). The use of multiple-try method and local optimization in Metropolis sampling. *J. Amer. Statist. Assoc.*, **95**, 121-134.

Liu, J.S., Liang, F., and Wong, W.H. (2001). A theory for dynamic weighting in Monte Carlo computation. *J. Amer. Statist. Assoc.*, **96**, 561-573.

Liu, J.S., Neuwald, A.F., and Lawrence, C.E. (1995). Bayesian models for multiple local sequence alignment and its Gibbs sampling strategies. *J. Amer. Statist. Assoc.*, **90**, 1156-1170.

Liu, J.S. and Sabatti, C. (2000). The generalized Gibbs sampler and multigrid Monte Carlo for Bayesian computation. *Biometrika*, **87**, 353-369.

— (1999). Markovian structures in biological sequence alignments, *J. Amer. Statist. Assoc.* **94**, 1-15.

Liu, J.S., Wong, W.H. and Kong, A. (1994). Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes. *Biometrika*, **81**, 27-40.

— (1995). Correlation structure and convergence rate of the Gibbs sampler with various scans. *J. Roy. Statist. Soc. B*, **57**, 157-169.

Liu, J.S. and Wu, Y.N. (1999). Parameter expansion for data augmentation. *J. Am. Statist. Assoc.*, **94**, 1264-74.

MacEachern, S.N., Clyde, M.A., and Liu, J.S. (1999). Sequential importance sampling for nonparametric Bayes models: the next generation. *Canadian J. Statist.*, **27**, 251-267.

McCue, L.A., Thompson, W., Carmack, C.S., Ryan, M.P., Liu, J.S., Derbyshire, V., Lawrence, C.E. (2001). Phylogenetic footprinting of transcription factor binding sites in proteobacterial genomes. *Nucl. Acids Res.*, **29**, 774-782.

Neuwald, A., Liu, J.S., and Lawrence, C.E. (1995). Gibbs motif sampling: detection of bacterial outer membrane protein repeats. *Protein Science*, **4**, 1618-1632.

Neuwald, A.F., Liu, J.S., Lipman, D.J., and Lawrence, C.E. (1997). Extracting protein alignment models from the sequence database. *Nucl. Acid Res.*, **25** No. 9, 1665-1677.

Neuwald, A.F., Kannan, N., Poleksic, A., and Liu, J.S. (2002). Canonical Ras, Rab, Rho and Ran GTPase structural features and Rans C-terminal and basic patch mechanisms. *Genom. Res.*, tentatively accepted.

F.P. Roth, J.D. Hughes, P.W. Estep, and G.M. Church (1998). Finding dna regulatory motifs within unaligned noncoding sequences clustered by whole-genome mrna quantitation. *Nat. Biotech.*, 16: 939-45.

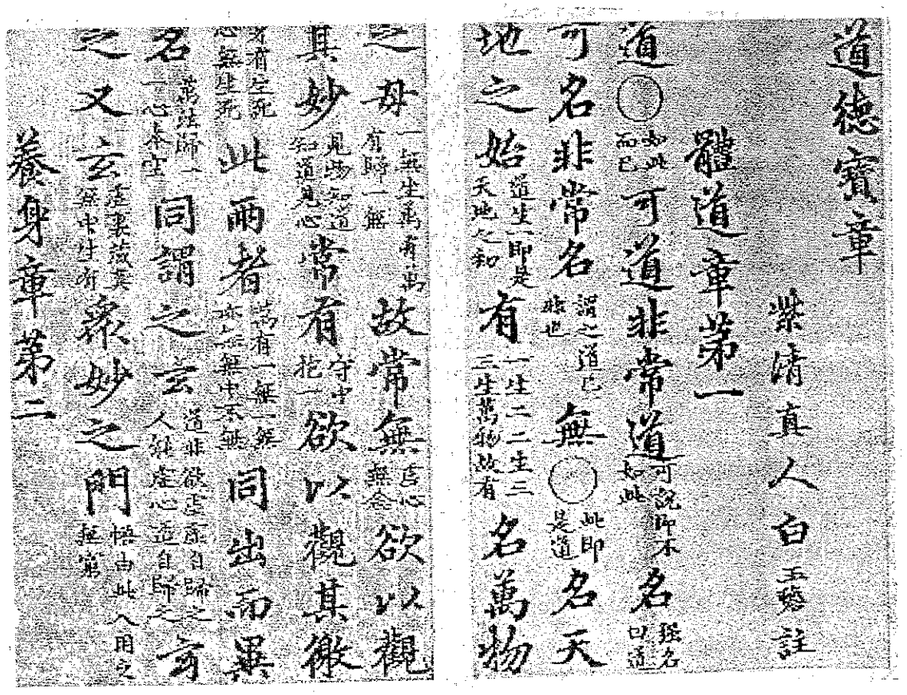
Storch, K.F., Lipan, O., Leykin, I., Viswanathan, N. Davis, F.C., Wong, W.H., and Weltz, C.J. (2002). Extensive and divergent circadian gene expression in liver and heart. *Nature*, 417, 78-83.

J. D. Thompson, D. G. Higgins, and Gibson T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, 22(22):4673–80.

W. W. Wasserman, M. Palumbo, W. Thompson, J. W. Fickett, and C. E. Lawrence (2000). Human-mouse genome comparisons to locate regulatory sites. *Nature Genetics*, 26(2):225–228.

Xu, Y. Xu, D. and Olman, V. (2002). A practical method for interpretation of threading scores: a applicaton of neural network. *Statistica Sinica* 12, 159-178.

Zhang, J.L. and Liu, J.S. (2002) A new sequential Importance sampling method and its application to the 2D HP model. *J. Chem. Phys.*, **117** 3492-8.



Controversial Statistical Issue - *Flexible Design*

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

In the recent decade there has been considerable interest in achieving greater flexibility for designing and conducting a clinical trial. Examples of flexible designs or adaptive designs, depending on authors' definition, are many. Most often known is the sample size re-estimation based on interim data information, whether for the superiority objective or for the non-inferiority objective or for both. Applications can be found in drug trials for many disease areas, for instance, pulmonary, gastro-intestinal, oncologic, cardio-renal, etc.

Other types of flexible (or adaptive) designs that we encountered include modifications of the study hypotheses, for example, dropping treatment arms based on information from inside or outside the trial, changing study objectives between non-inferiority and superiority, changing the study endpoint, changing the test statistic, etc.

We know that valid statistical inference is essential for evaluation of clinical trial results, which might involve multiple inferences. One goal of flexible designs is to allow for proper mid-trial design modifications or for full adaptation in designing without compromising on validity of statistical inference and interpretation and with careful planning. However, still debated is a question: should flexible (or adaptive) design be used routinely? should it never be done? or, should it be used carefully with external monitoring, such as, by an independent data monitoring board, or even a small group from the drug sponsor?

By Sue-Jane Wang, Ph.D. The Food & Drug Administration. Email: wangs@cder.fda.gov.
The views expressed in this article are not necessarily of the U.S. FDA.

Adaptive/Flexible Designs in Clinical Trials

Peter Bauer, Ph.D.

The application of statistical methods to the evaluations of clinical trials has been (and is) dominated by a rigid "a-priori planning" paradigma. The reason for this can be traced to classical frequentist methodology which in general asks for a detailed predefinition of the decision process concerning number and time points of analyses, as well as the decision rules to be applied. Such a pre-specification is required because data driven mid-trial design modifications may lead to a violation of basic statistical properties considered to be relevant, e.g., in testing the probability to perform a type I error may be increased or in estimation bias may arise. The rigidity of the pre-planning paradigma has been criticized from different ankles. The experimenter usually is not able to supply all the detailed information, which would be needed to plan the "optimal" design (which mostly depends on the unknown "true" state of nature). On the one hand, if he would know all about that, he presumably would not run a trial at all. On the other hand he often will not accept that he can not make use of what he "learns from experience" during the course of the trial to improve on possible misspecifications in the planning phase. Another criticism is more fundamental and is used as an argument to propose alternative types of statistical inference. So in Bayesian methods statistical inference does not depend on a (pre-specified) stopping rule to be applied during the trial but only on the posterior distribution evolving at any time point.

It is not the purpose of this comment to discuss these fundamental issues. In the following it is outlined, how an enormous amount of flexibility for mid-trial design adaptations can be achieved within the context of frequentist inference and which price has to be paid for.

The basic idea behind is to partition a trial into different stages and calculate separate test statistics from the disjoint stages, which then are combined in a suitable pre-defined way. An example for a partitioning is group sequential designs where the analyses are performed after (pre-defined) increments in sample size (or information) have been achieved respectively. Whereas in group sequential designs the test statistics at the analyses times are calculated in the conventional way from the pooled data collected so far, in adaptive (flexible) designs separate test statistics are calculated for the incremental information cumulated since the previous analysis. Decisions are derived from a pre-defined combination of these separate test statistics over stages. As a simple example let us consider two-stage designs. Here stage wise (one-sided) p-values can be combined by a suitable level α combination test (Bauer, 1989), e.g., by Fisher's product criterion with combination function $p_1 \bullet p_2$ (Bauer& Köhne, 1994), or by the inverse normal method using $1/\sqrt{2})(\Phi^{-1}(1-p_1)+\Phi^{-1}(1-p_2)$ (Banik et al., 1996; Lehmacher & Wassmer, 1999) or its weighted extension applying $w_1\Phi^{-1}(1-p_1)+w_2\Phi^{-1}(1-p_2)$, $w_1^2+w_2^2=1$, w_1, w_2 prefixed (Lehmacher & Wassmer, 1999; Cui et al., 1999). The crucial point behind is that under the null hypothesis the distribution of the future stage-wise test statistics under very general conditions preserve their statistical properties even if the stages have been adapted immediately before they have been performed (Bauer et al. 2001). Note that all the information collected in the past may be used for the adaptation. To see that let us look at the simple, widely discussed example of mid-trial sample size reassessment when testing the mean of a normal distribution with known variance. In the interim analysis an experimenter may take the estimated effect for recalculating the sample size. Given that the null hypothesis is true, the p-value of the forthcoming stage (calculated in a sample of newly recruited patient) should have a uniform distribution on [0,1] independent from what has been observed previously at the first stage. Let z_1 be the standardized test statistics at the first stage, z_2' be the standardized test statistic calculated from the second stage sample with adapted sample size n_2'. Then, the use of the weighted inverse normal combination function with pre-fixed weights $w_1\Phi^{-1}(1-p_1)+w_2\Phi^{-1}(1-p_2) = w_1z_1+w_2z_2'$ in the final analysis will guarantee control of the type I error rate irrespective of the sample size reassessment driven by the effect size observed in the interim analysis. It is straightforward how early stopping rules for the first stage test statistics z_1 can be incorporated into the concept of combination tests. Here an important relationship to group sequential designs can be seen: If $w_1=\sqrt{(n_1/(n_1+n_2))}$, $w_2=\sqrt{(n_2/(n_1+n_2))}$, where n_1 and n_2 are the pre-planned sample sizes for the two stages, and no sample size reassessment is performed in the interim analysis then the weighted inverse normal combination function is identical to the classical pooled test statistic used in group sequential designs. Hence group sequential trials are a special case of adaptive (flexible) multistage designs and the whole methodology to derive critical boundaries for group sequential designs can be applied. However, if adaptations are performed, the test statistics in general differ from the sufficient test statistics in the pooled sample. This is the price to be paid for adaptations and has been criticised since the decisions of the adaptive test may differ from those based on classical test statistics.

The whole concept is closely related to that of the conditional error function (e.g., Proschan & Hunsberger, 1995), which can be looked at as in terms of adaptive combination test and vice versa (Posch & Bauer, 1999). The conditional error function at any time during the trial is defined as the probability to perform a type I error in the remainder of the design, given the results observed up to now. To see the relationship we consider the simple example of Fisher's product criterion in two-stage designs: Let $p_1p_2 \le c_\alpha$ be the

critical region in the final analysis. Assume that we deal with continuous test statistics and p_2 under the null hypothesis is uniformly distributed. Then rejection in the final analysis occurs if p_2 does not exceed the critical limit c_α/p_1 which is the conditional error probability given the observed p_1. This opens a further dimension of flexibility: Instead of spending the type I error in the pre-planned way at any time it can later on be spent in a adapted design which, given the data observed, does never lead to a higher type I error than the original design (Müller & Schäfer, 2001a). In the above example we can replace the forthcoming second stage by a design which for any z_1 applies a "local" test with a significance level equal to the conditional error function c_α/p_1. (The concept may even be applied in a situation where no interim analysis has been pre-planned (Müller & Schäfer, 2001b) by calculating the conditional type I error rate at any time point.) This generalization allows to keep the number of planned interim analyses open for adaptation; e.g., leaving out the next planned interim analysis if, based on the current results, there is no reasonable chance for a rejection (Posch et al. 2003). Again this can be looked at in terms of combination tests by recursively applying and planning two-stage combination tests. The recursive principle under very general conditions allows the definition of p-values and confidence intervals (Brannath et al., 2002).

Since the principle of combining stage-wise p-values is very general and can be applied for different distributional scenarios, various other types of adaptations have been discussed: Selection of treatments (doses) to be skipped for the remainder because of safety or lack of efficacy issues, reallocation of sample size between treatments, choice of appropriate scores in the test statistics to be applied at further stages, selecting a subset of multiple endpoints to be used further on, changing the order or the weights in a set of multiple endpoints and switching goals from non-inferiority to superiority testing (for a list of references see Posch et al., 2003). The key issue in most of these applications is that we are not any more dealing with a single null hypothesis as in sample size reassessment but with intersections of null hypotheses, which may vary between stages. Hence problems of interpretation will arise if the intersection of different null hypotheses tested at the two stages is rejected by the combination test. It has been shown how methods of multiple comparisons can be incorporated but this will complicate inference (Bauer & Kieser, 1999; Kieser et al. 1999; Hommel, 2001).

A few basics have to be stated when considering the value of adaptive (flexible) designs in applications. First of all the method is simple and using the appropriate combination function it coincides with group sequential tests, provided that no adaptations have been performed during the trial. As a price for performed adaptations in general test statistics have to be used which do not rely on the sufficient statistics in the cumulated sample. Constructing suitable point estimates and confidence intervals is already a fairly involved problem for classical sequential designs. The lack of a full pre-specification of the adaptation rule in the sample space or the adaptive choice of hypotheses introduce further complications. Dealing only with sample size adaptations a variety of other methods has been proposed so that it should be always carefully considered if the combination principle should be used for that purpose. Alternative methods include overpowering group sequential designs with the option of early rejections in case of clear effects (Jennison & Turnbull, 2003). However, in order to avoid overly large average sample sizes when being close to the null hypothesis stopping for futility and its impact on the development processes of new treatments in medicine will then be an important issue. The use of the observed effect size for sample size reassessment will lead to strongly varying and large average sample sizes. Generally, if the experimenter can lay down an adaptation rule a-priori, which he is going to strictly adhere to, then he may find methods for the analysis which are tailored to that particular adaptation rule and therefore will improve on the general combination principle (Tsiatis & Mehta, 2002)

the latter covering a much wider field of measurable adaptation rules which need not to be specified in advance (Brannath et al, 2002). But this is not surprising. Hence this general principle of adaptive (flexible) designs will have its merits particularly when not all potential mid-trial design modification can be laid down and modeled statistically a-priori. But this has to be expected in many experimental situations, where characteristically not all aspects of the decision process with regard to benefit, risks and costs can be foreseen in advance.

Reference

1. Banik, N., Köhne, K., Bauer, P. (1996). On the power of Fisher´s combination test for two stage sampling in the presence of nuisance parameters. *Biometrical Journal*, 38:25-37.
2. Bauer, P. (1989). Multistage testing with adaptive designs (with discussion).. *Biometrie und Informatik in Medizin und Biologie*. **20**, 130-148.
3. Bauer, P., Brannath, W., Posch, M. (2001). Flexible two stage designs: An overview. *Methods of Information in Medicine* **40**, 117-121.
4. Bauer, P., Kieser, M. (1999). Combining different phases in the development of medical treatments within a single trial. *Statistics in Medicine* **18**, 1833-1848.
5. Bauer, P., Köhne, K. (1994). Evaluation of experiments with adaptive interim analyses. *Biometrics 50*, 1029-1041.
6. Brannath, W., Posch, M., Bauer, P. (2002). Recursive combination tests. *Journal of The American Statistical Association*, **97**:236-244.
7. Cui, L., Hung, H.M.J., Wang, S. (1999). Modification of sample size in group sequential trials. *Biometrics* **55**, 321-324.
8. Hommel, G. (2001). Adaptive modifications of hypotheses after an interim analysis. *Biometrical Journal* **43**, 581-589.
9. Jennison, C., Turnbull, B.W. (2003). Mid-course sample size modification in clinical trial. Unpublished.
10. Kieser, M., Bauer, P., and Lehmacher, W. (1999). Inference on multiple endpoints in clinical trials with adaptive interim analyses. *Biometrical Journal* **41**, 261-277
11. Lehmacher, W., Wassmer, G. (1999). Adaptive sample size calculations in group sequential trials. *Biometrics* 55, 1286-1290.
12. Müller, H.H., Schäfer, H. (2001a). Adaptive group sequential designs for clinical trials: Combining the advantages of adaptive and of classical group sequential approaches. *Biometrics* **57**, 886-891.
13. Müller, H.H., Schäfer, H. (2001b). Changing a design during the course of an experiment. Unpublished.
14. Posch, M., Bauer, P.(1999). Adaptive two stage designs and the conditional error function. *Biometrical Journal* **41**, 689-696.
15. Posch, M., Bauer, P., Brannath , W. (2003). Issues in designing flexible trials. *Statistics in Medicine*. To appear.
16. Proschan, M. A., Hunsberger, S. A. (1995). Designed extension of studies based on conditional power. *Biometrics* **51**, 1315-1324.
17. Tsiatis, A.A., Mehta, C. (2002). On the inefficiency of the adaptive design for monitoring clinical trials. Paper presented at the Harvard/Schering-Plogh Symposium, Harvard School of Public Health. Submitted.

Peter Bauer, Ph.D.
Department of Medical Statistics, University of Vienna, Vienna, Austria
Email: peter_bauer@univie.ac.at

Adaptive Designs for Flexibility not Midcourse Change

Weichung J. Shih, Ph.D.

The field of clinical trials prior to 1960's was said to be a statistically undisciplined 'wild west'. Over the many decades we statisticians have championed the need of rigorous protocols and data analysis plans. For studies to be statistically sound, we have advocated pre-specifying study objectives, sample size/power, primary endpoints, and other design specs, then to stick to the specs and analyze data as planned. Making data-driven changes after the study started could be costly in terms of rejection / acceptance of the study results by speculators. Have we become petrified and can we make reasonable changes during a study? Many turn to 'adaptive designs' for answers. See (Shih, 2001) and (Wittes, 2002) for recent overviews.

For those who worry about study integrity due to changes, the good news is that 'adaptive design' is really not about making midcourse *changes* of a clinical trial. It is about *flexibility*. For those who seek relief from rigidity, the bad news is that the flexibility still comes from careful planning. Yes, by definition, adaptive designs still requires us to plan. Let me explain.

In a broader sense, any sequential design is adaptive, since the future course of a study is dependent on interim outcomes. But the specific adaptive design many people are discussing these days refers to the following characteristics: (a) an interim analysis is conducted at time t_i with n_i patients; (b) use the results of the n_i patients to decide the future course of the study; (c) the possible future course may be termination for either early efficacy or futility, or continuation with additional n_{i+1} patients to be recruited for another r_{i+1} units of time and cumulative data to be analyzed at the next review time t_{i+1}; (d) while continuing the study, may test another hypothesis (such as non-inferiority as opposed to superiority), another endpoint (such as overall survival as opposed to disease-free survival), etc.; (e) in doing so, satisfy certain operating requirements such as controlling the overall type-I error rate, securing adequate power (conditional power or overall power), and minimizing potential bias from looking at the interim data.

Notice that in the above setting, there is not a pre-specified 'sample size' n (or number of events) before the study starts, as in the traditional way of designing a study (including the group sequential designs.) Here, what one needs is the n_1 for the first analysis, not a total n. Of course, one can use an initially given sample size n to gauge the first analysis (i.e., let n_1 = a fraction of n), but this is mainly for convenience, not a statistical necessity. Hence, adaptive designs need not be regarded as 'mid-course change' in sample size (or total information, in general), if you do not give this nuisance n. On the other hand, the (c) and (d) parts involve more decision rules (than just the termination or continuation) based on the results of interim data, and the (e) part still requires constraints as in the traditional designs. See (Want et al., 2001) for selection of superiority/non-inferiority objectives and (Shih et al., 2003) for extension to primary endpoint from intermediate endpoint, after interim analyses. Again, these after-interim- analysis actions do not need to be regarded as 'mid-course changes', since they are a part of the flexible plan.

Some have raised the question of statistical efficiency relative to the fixed sample size design. To me, this is a non-issue, because the fixed sample size design does not allow the study itself to direct its future, but rely on some pre-trial assumptions on recruitment rate, effect size and variability without checking their reality for the current trial. Instead, we should think hard on how to determine an appropriate n_1 to do the first analysis. For example, for methods that use interim analysis to find an estimate of the mean treatment effect while assuming a constant variance (e.g., (Proshan and Hunsberger, 1995) and (Li et al., 2002)), we should seek n_1 so that

the variance estimate is stable. We also should argue the relevance of conditional versus unconditional power when decision is based on interim data. Finally, I think efforts are still needed to convince regulatory or grant authorities of the advantages of flexibility offered by adaptive designs, including having no budgetary-based sample size or study duration in the beginning of the trial but providing the timing of the first interim analysis and delineating possible decision rules for later actions.

Reference

1. Li G, Shih WJ, Xie T, Lu J (2002). A sample size adjustment procedure for clinical trials based on conditional power. *Biostatistics* **3**: 277-287.
2. Proschan M, Hunsberger S (1995). Designed extension of studies based on conditional power.*Biometrics* **51**:1315-1324.
3. Shih, WJ (2001). Sample Size Re-estimation – Journey for a Decade. *Statistics in Medicine* **20**: 515-518.
4. Shih WJ, Ouyang P, Quan H, Lin Y, Michiels B, Bijnens L (2003). Controlling type-I error rate for fast track drug development programs. *Statistics in Medicine*, to appear.
5. Wang SJ, Hung HMJ, Tsong Y, Cui L (2001). Group sequential test strategies for superiority and non-inferiority hypotheses in active controlled clinical trials. *Statistics in Medicine* **20**: 1903-1912.
6. Wittes J (2002). On changing a long-term clinical trial midstream. *Statistics in Medicine* **27**: 2789-2795.

Weichung J. Shih, Ph.D. is a Professor of Statistics Department, University of Medicine & Dentistry of New Jersey, NJ, USA
Email: shihwj@umdnj.edu

Adaptive Design for Drug Development

H.M. James Hung, Ph.D.

In the past decade, there has been increasing interest in developing adaptive designs applicable to clinical trials for drug development. Literature is abundant and growing fast on adaptive multi-stage designs or adaptive sequential designs with or without interim analysis that may lead to an early decision for rejection or acceptance of a clinical or statistical hypothesis. A partial list of references is provided at the end of this article.

For decades, classical fixed information designs have been a gold standard for clinical experiments. There are many reasons behind that. The designs will force the researchers to think ahead, gather prior knowledge relevant to the study as much as possible, and anticipate the unexpected in trial planning. The statistics have several attractive statistical optimality properties under this type of designs, such as, sufficiency, efficiency, unbiasedness. Once the protocol is finalized, any possible trial conduct change should be kept to the minimum and any trial modification would need to be extensively scrutinized. The trial is then run following the protocol very closely. Such designs have contributed a great deal to the public health research.

So why would adaptive designs be of interest? Let us explore what a classical fixed information design demands. For comparing the mean of the experimental treatment with the mean of the control treatment with respect to a response variable, the sample size is planned to detect a postulated treatment difference δ at a specified significance level α and with statistical power $1-\beta$. Given an assumed standard deviation σ of the response variable, the per-group sample size (assuming 1:1 ratio randomization) will be $n = 2[(z_\alpha + z_\beta)\sigma/\delta]^2$,

where z_α is the $(1-\alpha)$-th percentile of the standard normal distribution. Thus, basically, the sample size is calculated to detect the effect size δ/σ and the fixed information design calls for a conservative or "reasonable" estimate of the effect size prior to designing the trial. Overestimation of δ or underestimation of σ may result in severe underestimation of the sample size. In one type of clinical trial, δ may be a given fixed number. For instance, a drug has long been given for chronic use in treating a disease, such as hypertension. As a public health question, will such chronic use of this drug reduce morbidity or mortality in this disease population? In this case, there is much knowledge about the drug regarding safety, cost, treatment administration and one can choose δ to be a minimally worthwhile effect size. However, for new drug development, the available data (e.g., from small Phase I and II trials, external data, etc.) that could shed some light about the effect size and the safety experience for human use are all most often limited. The total development cost is materially under consideration. Therefore, almost surely, the clinical significance of the effect size can not be assessed at the initial design stage. In this case, the best one can do is to postulate δ on the basis of some "educated guess" of a range of plausible effect sizes and no such thing called the most conservative effect size is available at the design stage. The classical fixed-information design cannot serve well for drug development. Shouldn't the trial design allow a proper adjustment of sample size if the sample path observed at some interim times strongly suggests that the postulated effect size may have been substantially underestimated? This adjustment calls for adaptation of the trial design.

In chronic heart failure clinical trials, often two equally important endpoints, such as mortality alone and a composite of mortality and morbidity outcomes are to be tested. To address the issue of multiplicity adjustment for alpha level, a frequently used strategy is to designate one as primary and the other as secondary and to test the two endpoints using a gatekeeper decision rule. In essence, the entire alpha is then allocated to the primary endpoint. If the primary endpoint achieves statistical significance at the specified alpha level, then the secondary endpoint is tested at the same alpha level. If the primary endpoint fails to achieve significance, then no testing is performed to the secondary endpoint. With this strategy, which should be the primary endpoint is critical. If the selected primary endpoint is less sensitive than the secondary endpoint to detection of the treatment difference, then interpretation of statistical results can be very agonizing. We have seen at least two case studies in which the primary endpoint has a nonsignificant p-value but the secondary endpoint is at least nominally statistically significant. The gatekeeper fails to lead to a statistical conclusion and it is always very hard to make 'what if' arguments post hoc. Shouldn't the statistical design allow a proper correction at some interim time if the sample path suggests that the initial guess of what the primary endpoint should be is wrong? This correction calls for adaptation of the trial design.

Design and analysis of multiple dose trials are also difficult, particularly Phase III trials. Statistical significance is frequently interpreted under the consideration that some type of the overall alpha error as a result of multiple comparisons (e.g., between doses and/or between each dose and control) needs to be properly controlled. In addition, some dose(s) may have disappointed efficacy and some others may have substantial adverse effects. Unfortunately, the traditional Phase I and II trials don't always provide accurate projection of such undesirable effects. Shouldn't the statistical design allow the flexibility of dropping the doses that are highly likely to have such undesirable effects based on the interim sample data? This flexibility calls for adaptation of the trial design.

In active-controlled clinical trials without a placebo or no treatment arm, the efficacy of the experimental treatment may be established through two routes. One is showing that the experimental treatment is more effective (Superiority) than the active control comparator (given that active control is efficacious). Another

is showing that the experimental treatment is not much less effective (Non-inferiority) than the active control. Depending on the conjecture about the true state of nature, the sample size may initially be planned to test for superiority or for non-inferiority. If testing for these objectives may entail different sample sizes, shouldn't the trial design permit an interim adjustment of sample size if necessary? Such an adjustment calls for adaptation of the trial design.

The examples given above are just a few and there are many others not mentioned here. We statisticians need to ask a simple question. Is there a unique design that can serve well in most of the applications? As a scientist, we ought not to trivialize the complexity of practicality. Practicality needs support from, should not be a hostage of, statistical doctrine. While the classical fixed information design continues to serve as a valuable benchmark, its inflexibility needs to be fairly challenged, especially for designing a Phase III clinical trial in drug development. Adaptive designs may offer reasonable approaches to balancing between order and chaos. The performances of adaptive designs need to be fairly explored, not necessarily by the traditionally used yardstick. After all, the long-term probability we statisticians are accustomed to may be relevant to the evaluation of the design properties but may be irrelevant to the success of an on-going individual trial.

Acknowledgments

This was supported by the RSR funds #96-010A, #99/00-008, and #02-06, awarded by the Center for Drug Evaluation and Research, Food and Drug Administration. Thanks are due to Drs. Lu Cui, John Lawrence, Sue-Jane Wang, Gordon Lan, Yi Tsong, Robert O'Neill, and Qin Liu for many helpful discussions.

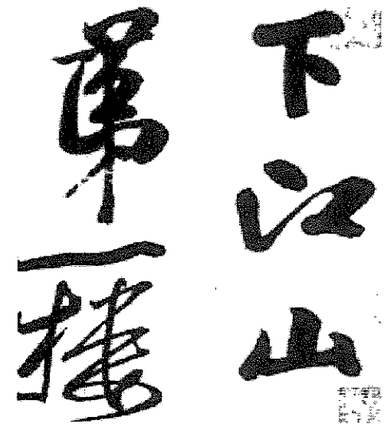
Disclaimer

The views expressed in this article are not necessarily of the U.S. Food and Drug Administration.

Reference

1. Bauer, P., Köhne, K. (1994). "Evaluations of experiments with adaptive interim analyses," *Biometrics* 50, 1029-1041.
2. Bauer, P., Röhmel, J. (1995). "An adaptive method for establishing a dose response relationship,: *Statistics in Medicine* 14, 1595-1607.
3. Bauer, P., Brannath, W., Posch, M. (2001). "Flexible two-stage designs: an overview," *Methods of Information in Medicine* 40, 117-121.
4. Berry, D.A. (2002). "Adaptive trials and Bayesian statistics in drug development," *Biopharmaceutical Report* 9, 1-7.
5. Brannath, W., Posch, M., Bauer, P. (2002). "Recursive combination tests," *Journal of the American Statistical Association* 97, 236-244.
6. Cui, L., Hung, H.M.J, Wang, S.J. (1999). "Modification of sample size in group sequential clinical trials," *Biometrics* 55, 321-324.
7. Cui, L. (2002)."Sample size re-estimation based on observed treatment difference," *Encyclopedia of Biopharmaceutical Statistics*, 2nd edition.
8. Fisher, L.D. (1998). "Self-designing clinical trials," *Statistics in Medicine* 17, 1551-1562.
9. Gould, A.L., Shih, W.J. (1992). "Sample size re-estimation without unblinding for normally distributed outcomes with unknown variance," *Communications in Statistics, Theory and Methods* A21, 2833-2853.
10. Gould, A.L., Shih, W.J. (1998). "Modifying the design of ongoing trials without unblinding," *Statistics in Medicine* 17, 89-100.
11. Hung, H.M.J., Cui, L., Wang, S.J., Lawrence, J. (2002). "Adaptive statistical inference following sample size modification based on interim review of effect size", 2002 *Proceedings of the American Statistical Association* (ASA), Biopharmaceutical Section [CD-ROM], Alexandria, VA: ASA
12. Jennison, C., Turnbull, B.W. (1999). "*Group sequential methods with applications to clinical trials*," Chapman and Hall/CRC Press, N.Y.
13. Jennison, C., Turnbull, B.W. (2001). "Mid-course sample size modification in clinical trials based on the observed treatment effect," Presented at 4th International Meeting on Statistical Methods in Biopharmacy, Paris.
14. Keiser, M., Bauer, P., Lehmacher, W. (1999). "Inference on multiple endpoints in clinical trials with adaptive interim analyses," *Biometrical Journal* 41, 261-277.
15. Lehmacher, W., Wassmer, G. (1999). "Adaptive sample size calculation in group sequential trials," *Biometrics* 55, 1286-1290.
16. Lan, K.K.G., Trost, D.C. (1997). "Estimation of parameters and sample size re-estimation," *Proceedings of the Biopharmaceutical Section*, American Statistical Association, 48-51.
17. Lawrence, J., Hung, H.M.J. (2002)."Estimation and confidence intervals after adjusting the maximum information," *Biometrical Journal* (to appear).
18. Li, G., Shih, W.J., Xie, T., Lu, J. (2002). "A sample size adjustment procedure for clinical trials based on conditional power," *Biostatistics* 3, 277-287.
19. Liu, Q., Chi, G. (2001). "On sample size and inference for two-stage adaptive designs," *Biometrics* 57, 172-177.
20. Liu, Q., Proschan, M.A., Pledger, G.W. (2002). "A unified theory of two-stage adaptive designs," *Journal of the American Statistical Association* (to appear).
21. Müller, H., Schäfer, H. (2001). "Adaptive group sequential designs in clinical trials: combining the advantages of adaptive and of classical group sequential approaches," *Biometrics* 37, 886-891.
22. Posch, M., Bauer, P. (1999)."Adaptive two stage designs and the conditional error function," *Biometrical Journal* 41, 689-696.
23. Proschran, M.A., Hunsberger, S.A. (1995). "Designed extension of studies based on conditional power," *Biometrics* 51, 1315-1324.
24. Shen, Y., Fisher, L.D. (1999). "Statistical inference for self-designing clinical trials with a one-sided hypothesis," *Biometrics* 55, 190-197.
25. Shih, W.J. (1992). "Sample size reestimation in clinical trials," In Peace, K. (ed.), *Biopharmaceutical Sequential Statistical Applications*, New York: Dekker, 285-301.
26. Shih, W.J. (2001). "Sample size re-estimation - journey for a decade," *Statistics in Medicine* 20, 515-518.
27. Tsiatis, A.A., Mehta, C.R. (2002)."On the inefficiency of the adaptive design for monitoring clinical trials," *Biometrika* (to appear).
28. Wang, S.J., Hung, H.M.J., Tsong, Y., Cui, L. (2001), "Group sequential test strategies for superiority and non-inferiority hypotheses in active controlled clinical trials," *Statistics in Medicine 20,* 1903-1912.
29. Wittes, J., Brittain, E. (1990). "The role of internal pilot studies in increasing the efficiency of clinical trials," *Statistics in Medicine* 9, 65-72.

H.M. James Hung, Ph.D.
Division of Biometrics I, Office of Biostatistics, OPaSS/CDER,
Food and Drug Administration,
Rockville, MD, U.S.A.
Email: hung@cder.fda.gov



Adaptive Designs in Clinical Trials

K.K. Gordon Lan, Ph.D.

As the methodology for clinical trials began to develop, the National Institutes of Health (NIH) sponsored a committee in the 1960s to develop guidelines for the conduct of clinical trials. This committee, chaired by Dr. Bernard Greenberg from the University of North Carolina, issued the Greenberg Report in 1967. The Greenberg Report endorses the concept of interim review of data by an independent Policy Advisory Board, which is now known as the Data and Safety Monitoring Board (DSMB) or the Data Monitoring Committee (DMC). For NIH sponsored trials, the DMC is generally an independent group of scientists not directly involved with the conduct of the trial. The DMC reviews the trial unblinded data periodically and then makes recommendations to NIH (or the sponsor) who would make the final decisions.

In order to handle clinical trial interim analyses appropriately, the subject matter demands two different concentrations – the biological mechanism and the application of suitable statistical procedures – be equally examined by experts from both the medical and statistical areas respectively. In most of the DMC meetings that I have attended, I found this "division of labor" to be true, that is, the medical DMC members (clinicians) were primarily concerned with the biological mechanism of the trial while the biostatisticians focused on the statistical issues involved. When bringing statistical procedures into clinical trials, the statisticians have the tendency to employ what we learned in graduate school, or what we read in the literature. However, we often forgot that those procedures were not motivated by clinical trials. For example, modification of the design of an experiment based on accrued data has been in practice for thousands of years in scientific research. Unfortunately, we have not spent enough effort in this aspect and demand many of the design parameters to be specified at the beginning of the study.

For a clinical trial sponsored by a pharmaceutical company, it is beneficial for the sponsor to have access to interim analyses results as soon as possible so that they can take appropriate actions – modify the design of the study for a more efficient evaluation, shift efforts to a back-up compound, begin preparing for manufacturing of the drug, or to discontinue the program. Unfortunately, the current practice does not allow the sponsor to have easy access to the interim results for the following reasons:

Many experienced DMC members are predominantly concerned with scientific consideration and they prefer the NIH model. I remember before the last DMC meeting that terminated BHAT (The Beta-Blocker Heart Attack Trial, sponsored by the NHLBI) in 1981, the DMC members were advised that the financial savings should not be an issue when it came to the decision for early termination of the study. As a habit, many DMC members for a company-sponsored trial may hesitate to communicate with the sponsors during interim analyses despite the potential for financial savings to the company.

The FDA is conservative in conducting interim analyses. In my opinion, their concerns are two-fold: the control of the alpha level and the bias the interim analyses may introduce. The former concern can be handled by employing appropriate statistical procedures; the latter, however, has no simple solution. No statistical procedures could possibly correct the bias caused by the practice of clinicians of the study when they are aware of or suspect the treatment differences.

A company-sponsored trial may be quite different from a NIH-sponsored trial that the NIH model has to be modified. Here is one crucial point why a completely independent DMC is not appropriate: The medical members of the DSMB are the first ones to obtain the information in a new treatment and this information is EXTREMELY valuable. As these DMC members are obviously the experts in their field of research, they could be members of another DMC for a similar trial or they could be conducting similar research in their own laboratories or clinical centers where the application of this information would be beneficial. The first hand information obtained at the DMC meetings may speed up their own research or motivate them to search for a different mechanism for the disease being studied. As the sponsor of a trial, a company deserves at least an equal right to access the information, and it may take an innovative approach to reach this goal.

Statistical research on adaptive design is just the first step to improve efficiency of drug development. Implementation of the research results to practice is the difficult part. To both ensure that the sponsor has access to all the crucial interim analyses information and to address the concerns of the FDA, I recommend that some selected company employees be allowed to attend the DMC closed sessions. Also, important information of development should be communicated to some designated upper managers of the sponsor. These employees and managers are required to go through a training course given by experienced experts in DMC operations, with input from the FDA.

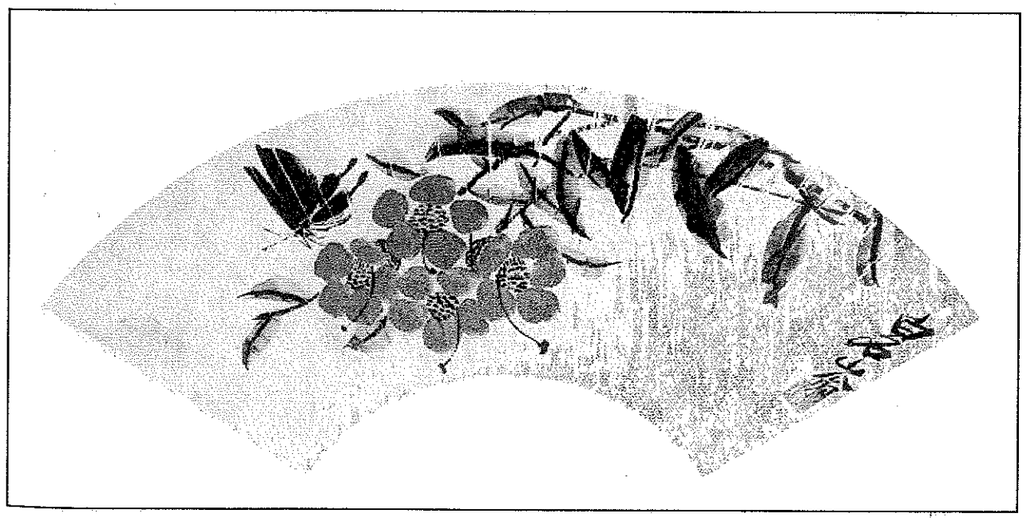
In addition, certain new SOP's will need to be developed to ensure that the sensitive information obtained during DMC meetings does not leak. Pharmaceutical companies should work with the FDA and the clinical trial community to develop such a process.

We all know that there are inside-traders in Wall Street. No one would imagine closing the Wall Street because of this unfair practice. Instead, people are working hard to improve the existing system. Why can't we do the same thing in drug development?

Reference

1. Lan, KKG ,"Problems and Issues in Adaptive Clinical Trial Design" , presented at ICSA Applied Statistics Symposium; Plymouth Meeting, Pennsylvania, June 8, 2002.

K. K. Gordon Lan, Ph.D. is a Senior Distinguished Specialist at Aventis Pharmaceuticals, Bridgewater, NJ, U,S.A.
Email: gordon.lan@aventis.com



Practical Aspects of Adaptive Designs in Pharmaceutical Research

Gernot Wassmer, Ph.D.
Reinhard Eisebitt, Ph.D.

In recent years several procedures were developed that can handle the issue of mid-trial design modifications in a clinical trial. These 'adaptive' or 'flexible' designs allow data-driven reassessments of essential characteristics of the study under protection of the overall type I error level. All information available in an interim analysis may be used for planning the subsequent stages of the trial. A key area is sample size re-estimation based on the observed effect or the variability of the endpoint. For example, based on conditional power arguments, it is possible to adjust the sample size in order to rescue a possibly underpowered trial. Adaptive designs, however, involve more general adaptations such as changing the number of interim analyses, changing the test statistic, dropping treatment arms, selecting endpoints.

The general theory rests upon the combination of the p-values for the separate stages of the trial through an appropriate combination function. This principle was first proposed by Bauer and Köhne (1994) in the context of two-stage test designs using Fisher's combination test. An appealing class of adaptive designs is given by the combination of the p-values through the weighted inverse normal method, because this turns out to represent a straightforward generalization of the classical group sequential test designs. Particularly, if no adaptations were performed during the course of the trial, the methods of data analyses are the same for both designs. This extension of the group sequential test procedure was described by Lehmacher and Wassmer (1999) in the context of multistage test designs. An analogous approach was independently proposed by Cui et al. (1999). Wassmer et al. (2001) describe statistical issues in planning and conducting such a trial. There is still little published work on controlled clinical trials with adaptive designs. Nevertheless, many pharmaceutical companies have conducted such trials in recent years in order to gain practical experience with this new class of statistical designs. We reviewed trials available to us in terms of actually performed changes in the design. In our internal files 40 interim analyses from 22 European studies with adaptive test plans were available. These studies were performed between 01/2000 and 06/2002. The results were as follows:

Clinical Phases - The majority of trials concerned the clinical Phases III (55% of the trials) and IV (41%), perhaps because these studies are usually more relevant and costly, thus making measures which maximize the probability of success of the trial more interesting in late clinical phases. Although Phase II research is the classical area where flexible designs may be used, only 4% of the studies were in Phase II. In Phase II trials, there is typically more uncertainty about the effects of the drug and the endpoints that characterize the drug effect(s). Adaptive designs may help explore the drug in Phase II and provide reliable effect estimates, which are needed for the planning of late Phase trials. However, some of the studies reported as belonging to Phase III had actually been a combination of Phase II (dose-finding) and Phase III. These studies had started comparing different dose levels of the drug to placebo or to an active control. In interim analyses, the dose with the best response was chosen and the sample size was adjusted appropriately. Finally, the studies proceeded until the effect of the test drug was proven in the confirmatory sense. Besides the flexibility in terms of sample size, which is obviously the most important advantage of adaptive designs, studies linking Phase II and Phase III have the potential that the evidence from Phase II can be used in the final confirmatory (Phase III) test statistic. That means data from Phase II and Phase III can be 'pooled' for the proof of effect. Hence, sample size can be saved.

Statistical Designs - Adaptive test plans were used for a wide range of statistical designs covering both superiority (78%) and non-inferiority trials (22%) with a single primary endpoint (78%), but also with multiple primary endpoints (4%) or combined endpoints (18%). Adaptive designs can help in dealing with the issue of changing the objective from non-inferiority to superiority.

Number of Interim Analyses - Very few study protocols planned only one interim analysis (4%). The vast majority of studies was planned with at least two interim analyses (two interim analyses: 50%, three: 14%, four: 32%). Clearly, the more interim analyses can be performed, the more the study will profit from the adaptive design. Of course, practicality demands a limitation of the number of interim analyses and too many changes may compromise the integrity of the study. While the adaptive methodology is developing to a mature level it may turn out that 2 or 3 interim analyses might represent an optimum.

Study Outcome - Theoretically, flexible designs allow for a wide range of adaptations such as changing or dropping of endpoints, changing the hypothesis, or dropping of treatment arms. However, the review shows that in practice adaptations other than sample size reassessments (38%) were exceptional. In 35% of the interim analyses the study was stopped without any adaptations having been performed, in 25% of cases no action at all was taken and the study was continued as planned until the next interim analysis, and in only 2% the design was changed.

In summary, our review reveals that (1) in practice the terms 'adaptive' and 'flexible'; (2) do not entail frequent design changes; (3) in about 10% the adaptive design can help to rescue the study; and (4) the involvement of a statistician and the availability of an appropriate statistical software is imperative, because maintaining the statistical integrity of the study (keeping the type I error) is more difficult in the 'flexible' case.

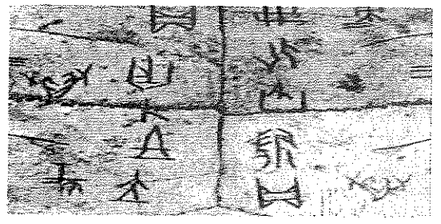
A false negative study outcome had been avoided through sample size adaptations or design changes in a number of studies, which otherwise would have been negative. Hence, adaptive designs were a useful tool for designing these trials. This methodology is rapidly developing and will become soon a standard tool in the statistician's repertoire.

Reference

1. Bauer P, Köhne K. (1994): Evaluation of experiments with adaptive interim analyses. *Biometrics* 50: 1029-1041.
2. Cui L, Hung HMJ, Wang SJ. (1999): Modification of sample size in group sequential clinical trials. *Biometrics* 55: 853-857.
3. Lehmacher W, Wassmer G. (1999): Adaptive sample size calculations in group sequential trials. *Biometrics* 55: 1286-1290.
4. Wassmer G, Eisebitt R, Coburger S. (2001): Flexible interim analyses in clinical trials using multistage adaptive test designs. *Drug Inf. J.* 35: 1131-1146.
5. Wassmer G, Eisebitt R. (2002): ADDPLAN Adaptive Designs, Plans and Analyses. Statistical Software for the Design, Simulation and Analysis of Adaptive Clinical Trials. User's Guide, Release 1.0. ADDPLAN GmbH, Cologne, Germany. www.addplan.org

Gernot Wassmer Ph.D.
Department for Medical Statistics, Informatics, and Epidemiology, University of Cologne, Germany
Email:gernot.wassmer@medizin.uni.koeln.de

Reinhard Eisebitt, Ph.D.
ClinResearch GmbH, Cologne Germany



On the Superiority of Adaptive Designs

Qing Liu, Ph.D.
Gordon Pledger, Ph.D.

Abstract - Under a group sequential design, the conditional power at an interim analysis can be either too high or too low. By adjusting the sample size, the conditional power can be maintained at a desirable level, which improves the efficiency and effectiveness of the experiment.

Keywords - Adaptive design; Conditional power; Benefit-risk analysis; Sample size adjustment.

Introduction - The conditional power at an interim analysis is the probability of eventually rejecting the null hypothesis under the alternative hypothesis. There can be inadequate conditional power as a result of chance or having a smaller effect size than anticipated, irrespective of the overall planned power. For an experimenter who is aware of this fact, a low conditional power can be unacceptable, and therefore, he or she may be willing to increase the sample size even though this requires additional resources and time to complete the experiment.

Recently, this intuition has been contradicted by the literature. Tsiatis and Mehta (2002) and Jennison and Turnbull (2003) have argued against adaptive designs, claiming that adaptive designs are inefficient on the basis that the average sample size of the adaptive procedure of Cui, Hung and Wang (1999) is larger than that of group sequential designs. It is not at all clear, however, whether the lack of efficiency is inherent in all adaptive designs. To resolve the issue, it is important to examine closely the sample size adjustment procedures proposed in the literature. We found that in designs with asymmetrical stopping boundaries the naive estimates from interim analyses are biased towards zero. Thus, if the sample size adjustment uses the naive estimates instead of the effect size itself the adjusted sample size can be extreme. Because an average can be strongly affected by extreme values, it is not surprising that the average sample size of an adaptive design using the naive estimates can be larger than that of a group sequential design.

Despite this fact, it is still not clear whether adaptive designs with larger average sample sizes, and therefore, higher power, are inferior to group sequential designs.

Efficiency or Effectiveness - Tsiatis and Mehta (2002) presented a fundamentally flawed argument: the adaptive designs are inferior because they are inefficient. They developed a uniformly most powerful (UMP) criterion for group sequential designs, which extends the UMP criterion of the Neyman-Pearson theory for fixed sample size designs. We point out that the Neyman-Pearson theory is suitable only for data analysis as it answers the following question: given the data collected, what is the best way to analyze the data such that the power for rejecting the null hypothesis can be maximized? This is easily seen if the Neyman-Pearson theory is closely examined at its foundation: it assumes that the loss for incorrectly rejecting or accepting the null hypothesis is a constant, irrespective of the sample size. This makes sense only if the data are given for which different choices of the sample size are out of the question. It then remains to find a test procedure that maximizes the power. For designing an experiment, the theory cannot be applied. This is because with different choices of the sample size and power, the constant loss assumption of the Neyman-Pearson theory no longer holds. At the basic level the Neyman-Pearson theory does not answer the question of how high the power should be for a particular experiment. This has not been widely recognized even though the Neyman-Pearson fundamental lemma was developed seventy years ago. Because of this, perhaps, the efficiency criterion of the Neyman-Pearson theory is often incorrectly applied to designing experiments; see for example, Tsiatis and Mehta (2002) and Jennison and Turnbull (2003)

From an economic point of view, it is more relevant to select designs on the basis of a benefit-risk analysis. For sample size n, let the payoff for rejecting the null hypothesis be $S(n)$ and the cost of conducting the experiment be $C(n)$. Obviously, $C(n)$ is an increasing function of n. Unfortunately, the statistical literature has not fully recognized that for most applications an experiment completed sooner with a positive result will yield a higher payoff. For example, Thach and Fisher (2002) assumed that $S(n)$ is a constant, leading to the extreme conclusion that the pharmaceutical companies should aim for nearly *100%* power if they want to maximize the expected profits. To reflect reality, we assume $S(n)$ is a decreasing function of n. Consider a fixed sample size design. If the experiment rejects the null hypothesis, then the net payoff is $S(n)-C(n)$; otherwise, the net payoff is $-C(n)$. Thus, the expected net payoff, or benefit, denoted by $B(\delta, n)$, is $S(n)P(\delta,n)-C(n)$, where δ is the effect size and $P(\delta,n)$ is the power for rejecting the null hypothesis. The risk $R(\delta,n)$ is the cost of experiment $C(n)$. We now define effectiveness for fixed sample size designs: sample size n_2 is more effective than sample size n_1 if and only if $B(\delta, n_2) / R(\delta,n_2) > B(\delta, n_1) / R(\delta,n_1)$. Under convexity, which holds for most applications in drug development, an optimal sample size and power exist and can be easily obtained numerically. The benefit-risk approach can be extended to group sequential designs and adaptive designs in general.

Summary of Results - By the benefit-risk analysis, we find that when the sample size calculation is based on the method of Liu and Chi (2001) to maximize the benefit, adaptive designs can be more effective than group sequential designs. More specifically, adaptive designs can have a higher benefit and lower risk. On the other hand, adaptive designs with sample size adjustment based on the naive estimates of the effect size are less effective than group sequential designs. This is because the adjusted sample size tends to be much larger than the optimal sample size. In comparison, the sample size of a group sequential design is relatively closer to the optimal sample size.

Acknowledgement - The authors would like to thank Dr. Girish Aras for his constructive comments and suggestions.

Reference

1. Cui, L., Hung, J. H. M., and Wang, S. J. (1999). Modification of sample size in group sequential clinical trials. *Biometrics* 55, 853-857.
2. Jennison, C. and Turnbull, B. W. (2003). Mid-course sample size modification in clinical trials based on the observed treatment effect. to appear in Statistics in Medicine.
3. Liu, Q. and Chi, G. Y. H. (2001). On sample size and inference for two-stage adaptive designs. *Biometrics* 57, 172-177.
4. Thach, C. T. and Fisher, L. D. (2002). Self-designing two-stage trials to minimize expected costs. *Biometrics* 58, 432-438.
5. Tsiatis, A. A. and Mehta, C. (2003). On the inefficiency of the adaptive design for monitoring clinical trials.

Qing Liu, Ph.D. & Gordon Pledger, Ph.D.
Statistical Science
J&J Pharmaceutical Research and Development, L.L.C., Raritan, N.J. U.S.A.
Email: qliu@worldnet.att.net

Stay Tune For The Next Controversial Statistical Issue In July 2003:

Multiplicity In Hypothesis Tests

Interested contributors please contact the Special Topic Editor: Dr. Sue-Jane Wang (wangs@cder.fda.gov).

Regional Activities

Hailiang Yang, Ph.D.

Hong-Kong Area

International Conference on Applied Statistics, Actuarial Science and Financial Mathematics

The University of Hong Kong and The Hong Kong Polytechnic University is jointly organizing an International Conference on Applied Statistics, Actuarial Science and Financial Mathematics. This conference is to be held at the University of Hong Kong (December 17, 2002) and the Hong Kong Polytechnic University (December 18-19, 2002).

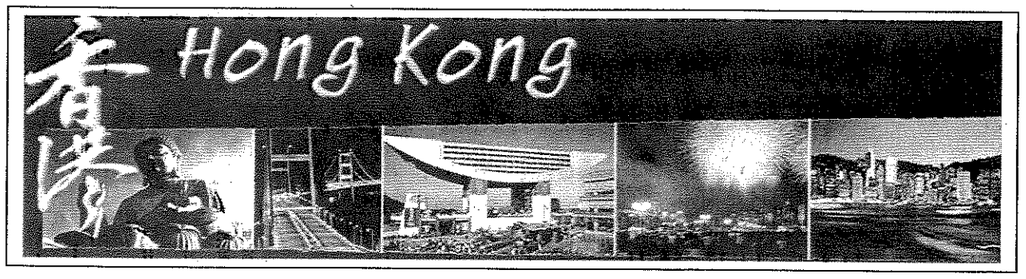
The chairmen of the organizing committee are Dr. K.W. Ng and Professor K.L. Teo and the chairmen of the scientific committee are Professors Elias Shiu and Howell Tong. The Society of Actuaries president, Harry Panjer, and the representative of the Institute of Actuaries (UK), Chris Daykin, will be addressing the conference. Invited speakers include leading experts and distinguished speakers in Statistics, Actuarial Science and Mathematical Finance. These consist of Shixue Cheng, A.H. El-Shaarawi, Jianqing Fan, Hans Gerber, Marc Goovaerts, Tze Leung Lai, David X. Li, Shige Peng, Howard Waters and Jiongmin Yong. There will be about 180 people attending the conference, about 70 of whom will come from Mainland China. There will be over 90 talks. For further information about the conference, please look at the conference web page on: http://web.hku.hk/~icaaf/

Workshop on Probability with Applications to Finance and Insurance

The University of Hong Kong held a workshop on probability with applications to finance and insurance from July 15-17, 2002. There were 32 invited talks in both theoretical and applied probability during the workshop. The workshop was initiated by Professor Tze-Leung Lai. Prof. Lai visited the University of Hong Kong as C.V. Starr Professor from June to July 2002.

Research Awards - Professor Howell Tong received the Distinguished Research Achievement Award from the University of Hong Kong in 2002. Professor W.K. Li received the Outstanding Researcher Award from the University of Hong Kong in 2002, and Dr. S.M.S. Lee received the Outstanding Young Researcher Award from the University of Hong Kong in 2002. The award presentation ceremony was held at the University of Hong Kong on November 4, 2002. Congratulations to all of them.

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



大陆统计学出现上升趋势

尽管在中国大陆统计学人材严重断层，但近年来，统计学发展出现了上升趋势。

自 1998 年在大学本科的学科分类中，统计学被列为一级学科（与数学并列）以来，已陆续有 50 所左右的高等学校招收统计学的班级，而且设立统计学班级的高校还在增加。

今年夏天，北京大学光华管理学院成立了统计与经济计量系，这是大陆统计界的一大喜事。此外，还有北京大学成立了大陆的第一个彩票研究中心，中国科学院数学与系统科学研究院成立了不确定性决策研究中心，并即将成立生物信息学研究中心，北京航空航天大学成立了复杂数据分析研究中心，这些对大陆统计学来说都是可喜可贺的事。

关于学术活动，今年五月份，每年一度的京津地区青年概率统计学术会议在中国人民大学如期召开。九月份，四年一次的全国概率统计学术会议在吉林省东北师范大学举行，参加人数达到了 205 人，比前两次会议增加了 40%以上，这主要是由于有 70 多位研究生和博士后参加了会议。这次会议的另一个特点是有 5 名日本客人和 3 名海外华人与会。会议共有 187 个学术报告，其中包括 3 个大会报告和 14 个中会报告。此外，今年还有一些大陆学者参加本土以外的学术会议。例如，有 50 余人参加了在香港举行的应用统计、精算科学和金融数学国际会议，有 5 人参加了在韩国举行的东亚统计联合会议，2 人参加了在韩国举行的第四届亚洲统计计算会议，有 4 人参加了在加拿大举行的 IMS 第 65 次年会，2 人参加了在纽约举行的 2002 JSM 等。

李国英　耿直

English Translation on the Next Page

Regional Activities

Guoying Li & Geng Zhi

Mainland China Area

Statistics in Mainland China

Even though Mainland China has a big gap in statistical researchers and teachers over the past decades, however, statistics has been very active and actually has been gaining great moment in recent years.

Since 1998, Statistics has been listed as a First-Level-Subject among other major subjects such as Mathematics in college curriculums. Consequently, more than 50 universities have classes for statistics and this trend has been increasing steadily.

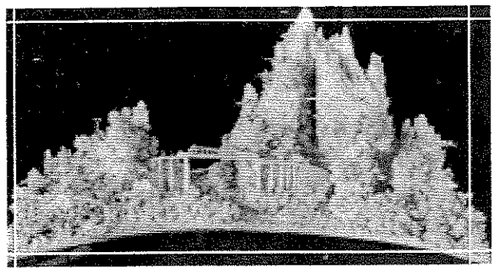
In the past summer, Guanghua School of Management of Peking University established the Department of Statistics and Econometrics. This has created great encouragement to the statisticians in Mainland China. In addition, several research centers closely related to statistics have also been established recently. For example, Peking University established the first Lottery Research Center in China. The Academy of Mathematics and System Sciences of Chinese Academy of Sciences has established a Stochastic Decision Research Center and a Bioinformatics Research Center. Beijing University of Aeronaustics and Astronautics has established a Research Center of Complex Data Analyses.

Statistical Conferences –

1. The annual Peking-Tianjing Area Probability and Statistics Conference for the Youth had been successfully held in May at the People's University of China.
2. In September, the quadrennial National Conference of Probability and Statistics had been successfully held at the Northeastern Normal University of Gilin Province. More than 200 people attended the conference. That was more than 40% increase compared with the past two conferences. This was primarily due to the participation of 70 graduate students and postdoctoral researchers. In addition to the native participants, five attendees were from Japan and three attendees were Chinese from abroad. A total of 187 papers were presented during the conference including three keynote addresses and 14 talks presented in median-size plenary sessions.
3. Many statistical researchers also attended conferences in other areas of the world. For example, about fifty researchers attended the International Conference on Applied Statistics, Actuarial Science and Financial Mathematics in Hong-Kong. Five researchers attended the East Asia Statistical Conference in South Korea. Two researchers attended the Fourth Asia Computational Statistics Conference in South Korea. Four researchers attended the 65th Annual IMS Conference in Canada. Two researchers attended the 2002 JSM in New York City.

Guoying Li is Professor of the Laboratories of Statistical Science, Institute of Systems Science, Chinese Academy of Sciences.
Email: gyli@mail.iss.ac.cn

Geng Zhi is a Professor of Mathematics at the Peking University, Peking, China.
Email: zgeng@math.pku.edu.cn



Regional Activities

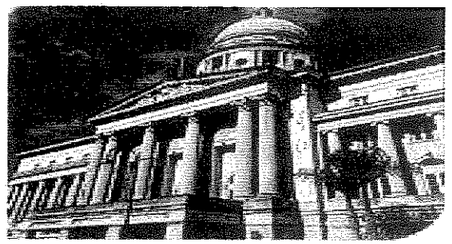
Singapore Area

June 1 – 30, 2003 – Statistical Methods in Microarray Analysis
Location: Singapore

This is a program of the Institute for Mathematical Sciences at the National University of Singapore. The Organizing Committee comprises: *Chair:* Terry Speed (University of California at Berkeley and Walter & Eliza Hall Institute of Medical Research, Australia); *Co-chairs:* Ming-Ying Leung (University of Texas at San Antonio) and Louxin Zhang (National University of Singapore); *Members:* Anthony Kuk (National University of Singapore), Art Owen (Stanford University), Sylvia Richardson (Imperial College) and Wing Hung Wong (Harvard University). More information is available at http://www.ims.nus.edu.sg/Programs/microarray/index.htm.

July 28 – August 31, 2003 – Stein's Method and Applications: a program in honor of Charles Stein
Location: Singapore

This is a program of the Institute for Mathematical Sciences at the National University of Singapore. The Organizing Committee comprises: *Co-chairs:* Andrew Barbour (University of Zürich) and Louis Chen (National University of Singapore); *Members:* Kwok-Pui Choi (National University of Singapore), Persi Diaconis (Stanford University), Larry Goldstein (University of Southern California) and Yosef Rinott (Hebrew University of Jerusalem).
More information is available at http://www.ims.nus.edu.sg/Programs/stein/index.htm



Call for Nomination
ICSA 2004 President & Board of Directors

The ICSA Nominating and Election Committee invites you to nominate candidates for the 2004 President-elect and five members of Board of Directors. Your input is important as the success of ICSA depends on strong leadership. We are looking forward to receiving nominees from you. Please check with nominees about their qualifications before you submit their names to Zhaohai Li, the Committee Chair. Nomination deadline is February 28, 2003. In your nomination, please include for each nominee: name, affiliation, address, and a paragraph of reasons for your nomination. You may nominate at most one candidate for the President-elect, and unlimited number of candidates for the board directors.

The candidate for the President-elect must
1. have been an active member for at least three years;
2. have held office of ICSA (e.g., director or committee member);
3. commit to chair board meetings during the year of presidency;
4. commit to attend at least one board meeting per year as president-elect and past president (exception can be made for candidates outside North America).

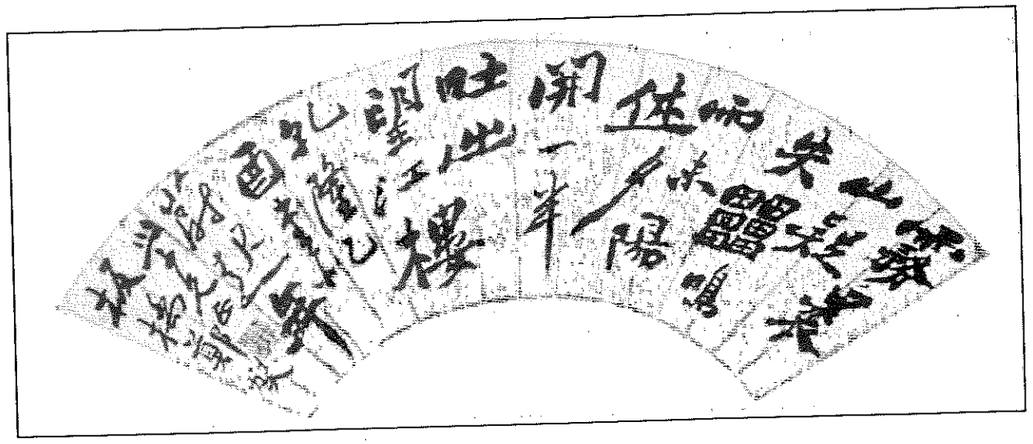
The candidate for Board Director must
1. be a current ICSA member;
2. commit to attend at least one board meeting per year, if resides in North America; attend at least two board meetings in three years if resides in Taiwan or Hong Kong; attend at least one board meeting in three years if resides in China;
3. Send a proxy, if unable to attend the board meeting.

Please submit your nomination to:

Professor Zhaohai Li
George Washington University
Department of Statistics
315 Funger Hall
Washington, DC 20052
Phone: 202-994-7844
Fax: 202-994-6917
Email: zli@research.circ.gwu.edu

Nominating and Election Committee:

Zhaohai Li, Ph.D. (Chair);
Committee Members: Jeff C. F. Wu, Ph.D., Naitee Ting, Ph.D., and Xuming He, Ph.D.



Call for Papers
The Sixth ICSA International Conference, 2004

July 21-23, 2004 - The Sixth ICSA International Conference, Singapore

The Sixth ICSA International Conference will take place at the National University of Singapore (NUS), Singapore, from July 21 to 23, 2004.

The Program Committee comprises: Zhiliang Ying (Columbia University) (Chair), Jia-Hua Chen (University of Waterloo), Louis H. Y. Chen (NUS), Paul Cheung (Chief Statistician of Singapore), Jianqing Fan (Chinese University of Hong Kong), Xuming He (University of Illinois at Urbana-Champaign), Dennis Lin (Pennsylvania State University), Xiaoli Meng (Harvard University), Kai W. Ng (University of Hong Kong), Qi-man Shao (NUS and University of Oregon), Young Truong (NUS and University of North Carolina at Chapel Hill), Qiwei Yao (London School of Economics), Yi-ching Yao (Institute of Statistical Science, Academia Sinica), Lincheng Zhao (University of Science and Technology of China).

The topics will include: probability theory, applied probability, statistical theory, official statistics, Bayesian computational statistics, longitudinal data analysis, econometrics, biostatistics, bioinformatics, genetical epidemiology, computational biology, semiparametric inference, time series, robust statistical methods, and mathematical finance.

The Local Organizing Committee will be chaired by Louis H. Y. Chen with Yiu-Man Chan as organizing secretary and Norizan binte Selamat as secretariat.

More information on the conference will be available on the web. For inquiries on the scientific program, please email Zhiliang Ying zying@stat.columbia.edu.

A panoramic view of Singapore's city skyline

ICSA 2003 APPLIED STATISTICS SYMPOSIUM

June 22-24, 2003 at San Diego, California, U.S.A.

Theme: Statistics in Bio-tech Research and Computing Intensive Methodologies

DATE: June 22 to 24, 2003. Short courses on Sunday, June 22, and technical sessions on Monday, June 23 and Tuesday, June 24.

LOCATION: University of San Diego. For local attractions, please visit the hotel website http://www.sandiego.org/whattodo.asp

CALL FOR PAPERS: The program committee invites you to submit statistical papers to be considered for presentation at the symposium. Abstracts are due March 31, 2003. Please submit abstracts to: Professor Gang Li, University of California at Los Angeles, email address: gangli@sunlab.ph.ucla.edu. The abstract should include the name, affiliation, mailing address, telephone number, fax number, and e-mail address of the author, and should not exceed 200 words. A template for the abstract can be downloaded from the ICSA website at http://www.icsa.org.

ICSA STUDENT AWARDS AND TRAVEL FELLOWSHIPS: The deadline is February 28, 2003 (see a separate page in this issue for detailed information). For further questions, please contact Professor Kung Jong Lui, San Diego State University, kjl@rohan.sdsu.edu

EXECUTIVE COMMITTEE
Nancy Lo (co-chair) (Nancy.Lo@NOAA.Gov) & Gang Li (co-chair) (gangli@sunlab.ph.ucla.edu).
Secretary: Alice Chu (alice.chu@pfizer.com)
Treasurer: Kathy Chi-Burris (kathy.chi-burris@pfizer.com)

Program Committee:
Gang Li (co-chair), Larry Shen (co-chair), Naihua Duan, Keh Shin Lii, Ying Lu, Kung Jong Lui, Edward Pun, Weng-Kee Wong, Eric Yan, Nancy Lo, and Joey C. D Lin.

Logistic Committee:
Nancy Lo (Chair), Alice Chu, Kathy Chi-Burris, William Yuan, David Shen, George Yu, John Lee, Thomas Lin, Kung Jong Lui, Xun Lin, Christina Show, Eric Yan, Joey C. D. Lin, Jenny Han, Edward Pun, Feng He, Ming Ji, Yan Wang, and Juangjuang Fan.

ICSA 2003 APPLIED STATISTICS SYMPOSIUM PRELIMINARY PROGRAM

- **Keynote Speakers (June 23-24, 2003):**
  Bradley Efron, Professor of Statistics and Biostatistics, Stanford University, President-Elect, ASA.
  http://www-stat.stanford.edu/people/faculty/efron.html
  George Tiao, W. Allen Wallis Professor of Econometrics and Statistics, University of Chicago, http://gsb.uchicago.edu/fac/george.tiao

- **Plenary Sessions (June 23-24, 2003):**
  Recent developments in nonparametric inferences with applications to biomedical studies and financial modeling, by Janqing Fan

  Current Statistical Issues in Clinical Trials for Drug Development, by Tze Leung Lai

- **Banquet Speaker: Arlene S. Ash, Boston University. (Lucky Star Chinese Sea Food Restaurant on June 23, Monday night.)**

- **Short Courses (Sunday, June 22, 2003, 9:00AM-5:00PM. See later pages for details)**

	Topic	Instructor
1	Practical Guidance of Generalized Linear Mixed Models	Charles E. McCulloch, University of California, San Francisco
2	Tutorial on Statistical Bioinformatics	Jun Liu, Harvard University
3	Cancer Trials for Practitioners – Experimental Design and Efficacy Analysis	Kao-Tai Tsai, Aventis Pharmaceuticals
4	Bootstrap Methods: A Guide for Practitioners	Michael R. Chernick, Novo Nordisk Pharmaceuticals
5	Active Controlled Clinical Trials	Yi Tsong & Sue-Jane Wang, FDA

- **Invited Sessions (June 23-24, 2003, not complete):**

	Session Topic	Organizer	Speakers
1	Statistical Applications in Business Research	Chih-Ling Tsai	Peter Lenk, University of Michigan Aaron Smith,UC-Davis Prasad A. Naik , UC-Davis
2	Non-Inferiority Testing in Active Controlled Clinical Trials	Yi Tsong	Steven Snapinn, Merck Ivan Chan, Merck Yong-Cheng Wang*, Gang Chen, and George Y. Chi, FDA
3	Statistics in Financial Econometrics	Jianqing Fan	Yongmiao Chen, Cornell University Per Mykland, University of Chicago Chunming Zhang, University of Wisconsin
4	Current Methodologies in Pharmaceutical Statistics	Kerry B. Hafner	Rafe Donahue, Glaxo Smith Kline Jimmy Wang, PRA International Shawn Yu, Takeda Pharmaceuticals, N.A.

5	Assessment of Measurement Agreement	Richard Runze Li	Lawrence I. K. Lin, Baxter Healthcare Corporation Runze Li and Mosuk Chow, Pennsylvania State University John Lu, National Institute of Standards and Technology
6	Recent Advances in Survival Analysis	Gang Li	Dorota Dabrowska, UCLA Jane-Ling Wang,UC Davis Mei-Cheng Wang, Johns Hopkins University
7	Data Mining in Chemistry and Chinese Medicine	Kai-Tai Fang	Yizhen Liang, Central South China University, China Kai-Tai Fang, Hong Kong Baptist University, Hong Kong, China Ricky N S Wong, Hong Kong Baptist University, Hong Kong, China
8	New Development in Medical Diagnostic and Screening Tests	Xiaohua Andrew Zhou	Margaret Pepe, University of Washington Vanja Dukic, University of Chicago Xiao-Hua Andrew Zhou, University of Washignton and VA Puget Sound Health Care System
9	Statistical Methods for AIDS Clinical Research	Hulin Wu	Hua Liang, Frontier Science & Technology Research Foundation Hulin Wu, Frontier Science & Technology Research Foundation Lang Wu, University of British Columbia
10	Design of Experiments	Ching-Shui Cheng	Weng Kee Wong, UCLA Hongquan Xu, UCLA Kenny Ye, SUNY at Stony Brook
11	Design and Analysis of Dose Response Studies	Naitee Ting	Douglas M Potter, University of Pittsburgh Cancer Institute Dr. Jason Hsu, Ohio State University Naitee Ting, Pfizer Global Research & Development
12	Statistical Applications in Accounting, Economics and Finance	Ruey S. Tsay	James Hamilton, UC San Diego Jimmy Ye, Baruch College Ruey S. Tsay, University of Chicago
13	Empirical Likelihood and Its Applications	Songxi Chen and Wei-Liem Loh	Jin Qin, Memorial Sloan-Kettering Cancer Center M. Tsao, University of Victoria, Canada Songxi Chen, National University of Singapore
14	Computing Intensive Methodologies in Bayesian Statistics	Minghui Chen	Jun Liu, Harvard University Steve MacEachern, Ohio State University Ming-Hui Chen, University of Connecticut
15	Bayesian Inference and Graphic Methods for Complex Data Analysis	Xiaoli Meng	TBA

16	New Development in Quality Improvements	Smiley W. Cheng	Gemai Chen and Lingyun Zhang, University of Calgary, Maria Tong, Sanford Papermate Corporation, Youn-Min Chou, U. of Texas at San Antonio
17	Statistical Methods for the Analysis of DNA and Tissue Microarray Data	Steve Horvath	Xinping Cui, UC Riverside Steve Horvath, UCLA Jim Veitch, Corimbia, Inc.
18	Markov Chain Monte Carlo and Its Applications	Dongchu Sun	TBA
19	Aspects of Clinical Trials	Grace Yang	Jian-Lun Xu and Richard Fagerstrom, Philip Prorok and Barnett Kramer George Yu, FDA Kai-Fun Yu and Aiui Liu, NIH
20	Bioengineering and Statistics	Nancy Lo	Yihua Zhao, UCSD John Shyy, UC Riverside Tzung Hsiai, USC
21	New Developments in Longitudinal Data Analysis	Annie Qu	Ying Qing Chen, UC Berkeley Wei Pan, University of Minnesota Annie Qu, Oregon State University Peter Song, University of Michigan
22	Adaptive Design for Clinical Trials	JianWen Cai	Yu Shen, MD Anderson Cancer Center Qing Liu, Johnson & Johnson Peter Thall, MD Anderson Cancer Center
23	Sequential Clinical Trials	Peng Huang	Aiyi Liu, NIH H.M. James Hung, FDA Peng Huang, Medical U. of South Carolina Sue Jane Wang, FDA
24	Statistics for Natural Resources	Nancy Lo	Steven K. Thompson, Penn State University Mark Maunder, IATTC Din Chen, International Pacific Halibut Commission
25	Design and Analysis of Cardiovascular Clinical Trials	H.M. James Hung	H.M. James Hung, CDER, FDA Gordon Lan, Aventis Pharmaceuticals Steve Snapinn, Merck Research Lab
26	Design and Analysis of Cancer Trials	Eric Yan	John O'quigley, UCSD Kao-Tai Tsai, Aventis Paul Bycott, Pfizer
27	Medical Statistics in Mainland China	Ji-Qian Fang	Zhi Geng; Xia Li and Zheng Guo; Cai Xia Li; Feng Chen
28	Intensive Computing in Genetic Applications	Frank Shen	TBA
29	Better Writing Skills for Success	W.K. Wong	Xioa Li Meng, Havard University Don Ylvisaker, UCLA
30	ASA SD Chapter Session	Duane Steffey	Joey Lin, SDSU Reena Deutsch, UCSD Mikhail Golovnya, Salford Systems

ICSA 2003 APPLIED STATISTICS SYMPOSIUM
SHORT COURSES

June 22 (Sunday), 2003

University of San Diego, 5998 Alcala Park, San Diego, CA, 92110-2492

- For extra information about the short course please contact: Ying Lu, Ph.D., Department of Radiology, UCSF, 415-502-4596 or e-mail to ying.lu@radiology.ucsf.edu, or visit ICSA web site: *http://www.icsa.org*
- Early registration date for short courses 04/30/2003. Courses may be canceled if there is no sufficient number of registered attendants by 05/15/2003.
- Registration to the Symposium program by 05/15/2003 is highly encouraged but is not required for short course participants.

Overview of Five Short Courses (9:00am – 5:00pm, June 22, 2003)

	Topic	Instructor
1	Practical Guidance of Generalized Linear Mixed Models	Professor Charles E. McCculloch, University of California, San Francisco.
2	Tutorial on Statistical Bioinformatics	Professor Jun Liu, Harvard University
3	Bootstrap Methods; A guide for Practitioners	Dr. Michael R. Chernick, Novo Nordisk Pharmaceuticals
4	Cancer Trials for Practitioners – Experimental Design and Efficacy Analysis.	Dr. Kao-Tai Tsai, Aventis Pharmaceuticals.
5	Active Controlled Clinical Trials	Drs. Yi Tsong and Sue-Jane Wang FDA

Course 1: An Introduction to Generalized Linear Mixed Models

Instructor: ***Charles E. McCulloch, Ph.D.***
Division of Biostatistics, University of California, San Francisco

Course Outline:

The class of generalized linear mixed models (GLMMs) is a broad class of statistical models generalizing both linear mixed models (LMMs) and generalized linear models (GLMs). As such it is capable of accommodating nonlinear responses, correlated data and non-normal distributions. This makes it quite useful in practice. For example, GLMMs give a natural way to specify a correlated data model for binary data.

After briefly introducing generalized linear models I will describe the extension to GLMMs. The focus in the course will be on approaches to modeling, methods of estimation and inference, and available software. The concepts will be illustrated on a number of case studies. I will briefly mention some areas of current research. An outline of the course follows

OUTLINE

I. Introduction
  A. Example: Chestnut blight
  B. Overview of course
II. Review: Linear Mixed Models and Generalized Linear Models
  A. Examples
  B. Fixed versus random factors
  C. Best Linear Unbiased Prediction
  D. Estimation and Tests in LMMs
  E. GLM Basics
  F. Estimation, tests, software
III. Generalized Linear Mixed Models
  A. Examples
  B. GLMM basics
  C. Best Prediction for GLMMs
  D. Inference for GLMMs
  E. Case studies
IV. Selected topics in current research
V. Summary and discussion

About the Instructor:

Dr. Charles E. McCulloch completed his Ph.D. in Statistics at Cornell in 1980. He is a Professor and Head of the Division of Biostatistics at the University of California-San Francisco Medical School and was previously Professor and the founding Chair of the Department of Statistical Science at Cornell University. He is an expert on the development and use of statistical methods for longitudinal data analysis and the analysis of hierarchical models. He conducts primary research in the areas of generalized linear mixed models and latent class models. He is the co-author of the Wiley texts Variance Components and Generalized, Linear, and Mixed Models and the forthcoming IMS monograph on generalized linear mixed models. He is a fellow of the American Statistical Association and an elected member of the International Statistical Institute. He was the primary lecturer for an NSF-CBMS Regional Research Conference in 1999 on the topic of generalized linear mixed models. He has over 20 years of statistical consulting and collaborative experience.

Course 2: Tutorial on Statistical Bioinformatics

Instructor: ***Jun Liu, Ph.D.,***
Department of Statistics, Harvard University

Course Outline:

A substantial core of computational biology (or bioinformatics) methods has been developed during the past two decades to meet the need of biological scientists for data storage, data retrieval, and data analysis. A main problem that motivated early research in computational biology is protein sequence analysis. Recently, because of the dramatic increase in many types of biological data due to the human genome project and other high-throughput projects, the scope of bioinformatics research has been extended to embrace diverse topics such as micro-array analysis, protein classification, regulatory motif analysis, RNA analysis, structural and functional predictions, gene prediction, etc.

This short course is intended to provide coverage of some key developments of bioinformatics in the past thirty years with an emphasis on topics of recent interest. Topics include: pair-wise sequence analysis, local alignment, dynamic programming, BLAST, multiple sequence alignment, Gibbs motif sampler, gene regulation, hidden Markov models, context-free grammars, protein structure analysis, comparative genomics, clustering methods for microarrays, phylogenetic trees, etc.

OUTLINE

Background. The databases of DNA and protein sequences contain millions of sequences, many completed genomes, and more are coming rapidly. DNA microarray data are being produced at a phenomenal speed. Protein arrays are being developed. High throughput structural data are being produced. Analysis of these data using bioinformatics

tools has played a key role in several recent advances in biomedicine.

Hypes. Many have recognized the importance of the development of methods for the analysis of these data: Major drug companies are forming large bioinformatics groups; computer science departments have made bioinformatics a major focus; many government funding agencies have solicited bioinformatics grant applications; even statisticians start to talk about gene expressions and gene regulations.

Challenges. The sheer amount and variety of the molecular biology data have already presented a major challenge to all quantitative researchers. A distinctive feature of these data, be they microarray images, DNA sequences or protein structures, is that there is a large body of biological knowledge associated with them. This makes standard data mining or statistical analysis tools less effective. Incorporating relevant scientific knowledge into the development of statistical or computational analysis tools is the key to success.

About the Instructor:

Dr. Jun Liu completed his Ph.D. in Statistics at the University of Chicago in 1991. He is a Professor of the Department of Statistics and Biostatistics at Harvard University. He is an expert on Markovian structure and its relation to genetics; Bayesian computation methods and their applications in material science, chemistry, and structural biology; nonlinear state-space models; digital signal restoration; target tracking; and gene regulation; sequence alignment; protein structure prediction, etc. Among many distinguished honors he received, he was the recipient of the COPSS award 2002, 2002 IMS Medallion Lecturer, and recipient of Mitchell Award from the International Society of Bayesian Analysis. He is an associate editor of *JASA*, *Statistica Sinica, and Biometrics*. He has a book recently published by Springer entitled *Monte Carlo Strategies in Scientific Computing*. He has published in both statistical and biological journals.

Course 3: Bootstrap Methods: A Guide for Practitioners

Instructor: ***Michael R. Chernick, Ph.D.,***
Novo Nordisk Pharmaceuticals

Course Outline:

This is a six hour tutorial on bootstrap methods. It is intended for practicing statisticians and others with interest in applying statistical methods. Students should have a basic introduction to statistical methods and should be familiar with concepts of estimation, hypothesis testing, regression and analysis of variance. High school algebra is assumed but no higher-level mathematics is required. Some familiarity with computer simulation would also be helpful. The student will be provided with historical background on bootstrap methods and will be given a formal introduction to these methods. Emphasis will be placed on the wide variety of applications of the techniques, the computer-intensive nature of implementation along with many examples and "real world" applications.

OUTLINE

A) *Introduction to Bootstrap Methods:* In this segment I will give a formal introduction to the bootstrap including the history of its development. The Monte Carlo approximation to the bootstrap estimate will be emphasized and demonstrated in the simplest cases.

B) *Wide Variety of Applications:* I will emphasize in this segment the minimal amount of assumptions required for the bootstrap to work and present the basic idea of letting the data speak for itself. I will discuss the diverse areas of application and present a large number of problems that can be addressed using bootstrap methods.

C) *Specific Applications:* In this segment I will discuss bootstrap applications to estimating bias, constructing confidence intervals and testing hypotheses. The one-to-one relationship between confidence intervals and hypothesis tests will be emphasized. Various bootstrap confidence interval procedures will be discussed and compared including the percentile method and Efron's BCa method. The guidance given by Carpenter and Bithell will be discussed as well as special considerations for hypothesis testing (ideas of Hall and Wilson).

D) *Examples:* In this segment I will show examples from my work in the medical device industry and from my consulting experience. These specific examples are i) applications of the bootstrap to do p-value adjustment in multiple testing situations, ii) confidence intervals for statistical process capability indices and iii) an example to test bioequivalence. In addition I provide some counterexamples that show special situations where the bootstrap idea fails and approaches to get around these problems

E) *Recent Advances:* This final segment will include some interesting work subsequent to the publication of my 1999 book including bootstrap estimation of individual bioequivalence and Kerr's application of the bootstrap in analysis of microarray data. The ICSA has contributed to the bootstrap literature with many pertinent articles in Statistica Sinica (12 articles sited in my book). Also there are 5 articles in the Chinese Journal of Applied Probability and Statistics that I cite in my book. I will highlight some of this work.

About the Instructor:

Michael Chernick is Assistant Director of Biostatistics at Novo Nordisk Pharmaceuticals, Inc. He has a Ph.D. in statistics from Stanford University. He has spent the past 22 years as a research statistician at the Oak Ridge National Laboratory and as an industrial statistician in the aerospace, insurance, medical device and pharmaceutical industries. He has over 25 publications in the leading probability and statistical journals and has taught elementary and advanced statistical courses at various universities in southern California. Dr. Chernick has applied resampling methods in his work since 1983. His textbook "Bootstrap Methods: A Guide for Practitioners" was published by John Wiley and Sons, Inc. in 1999 and is the basis for his lectures in the resampling course. Dr. Chernick is a Fellow of the American Statistical Association, and a member of the Institute of Mathematical Statistics, the Society for Industrial and Applied Mathematics, the International Biometrics Society and the Society for Clinical Trials. He is a past President of the Southern California Chapter of the American Statistical Association and is currently the elected chair for the Sports Statistics Section Program for the Joint Statistical Meetings in 2002. In 1983, Dr. Chernick was awarded the Jacob Wolfowitz Prize for the best theoretical paper in the American Journal of Mathematical and Management Sciences among the articles in the 1982 volume. Dr. Chernick has taught short courses on bootstrap methods at the JSM continuing education program in 2000, at the ASA traveling course series and as part of a course on resampling for the Institute for Professional Education. He has a second book Introductory Biostatistics for the Health Sciences that he is co-authoring with Professor Robert Friis (expected to be published by Wiley in late 2002 or early 2003).

Course 4: Cancer Trials for Practitioners – Experimental Design, and Efficacy Analysis.

Instructor: ***Kao-Tai Tsai, Ph.D.,*** Aventis Pharmaceuticals

Course Outline:

Cancer research is an important and special area in clinical trials. Therefore, the conducts of cancer trials are somewhat different from the clinical trials of other therapeutic areas. In this course, the trial designs and analyses from Phase 1 to Phase 4 will be discussed. The discussions will include the choice of proper designs, the selection of endpoints for various cancer types, sample size estimations, quality of life, and regulatory considerations. The focus of this course will be on the broad practical aspects of cancer trials concerning clinicians and statisticians instead of the mathematical derivations of procedures.

About the Instructor:

Kao-Tai Tsai, Ph.D. is a Senior Manager of Biostatistics at Aventis Pharmaceuticals in New Jersey. He had worked at AT&T Bell Laboratories previously and has been with pharmaceutical industry since 1989 working at the U.S. FDA, Schering-Plough Research Institute, Organon Pharmaceuticals, and most recently at Aventis Pharmaceuticals.

Course 5: Active Control Clinical Trials

Instructors: ***Sue-Jane Wang[1], Ph.D. and Yi Tsong[2], Ph.D.***

[1]Division of Biometrics II, [2]Quantitative Methods and Research Staff, Office of Biostatistics, Center of Drug Evaluation and Research, Food and Drug Administration

Course Outline:

In an active controlled non-inferiority trial without a placebo arm, there are many statistical issues regarding to objective, design, analysis and interpretation. The morning sessions will cover the general issues on two arms active control non-inferiority clinical trial without a placebo arm that it is often not entirely clear what the primary objective is. Is it to demonstrate that the experimental treatment preserves at least some fraction of the effect of the active control, or that the treatment is not much less effective than the active control, or that the treatment is efficacious? The active control effect is a parameter, the value of which is unknown. To test the hypothesis of effect preservation, the classical confidence interval approach requires specification of a non-inferiority margin, which is a function of the unknown active control effect. When the margin is estimated, it is also not clear what is the relevant type I error of making a false assertion about preservation of the active control effect. Two types of statistical methods for non-inferiority testing are synthesized test method and confidence interval method. The utility and pitfall of these statistical methods will be discussed. The TACT (Two-stage Active Control Testing) method will be introduced as a possible compromise between the two methods. The morning sessions will also explore what percent level of the control effect needs to be preserved so that through non-inferiority testing of effect retention one can assert the treatment efficacy within a desired level of the error rate.

The afternoon sessions will cover the special issues of active control clinical trials. The topics to be covered include the following,

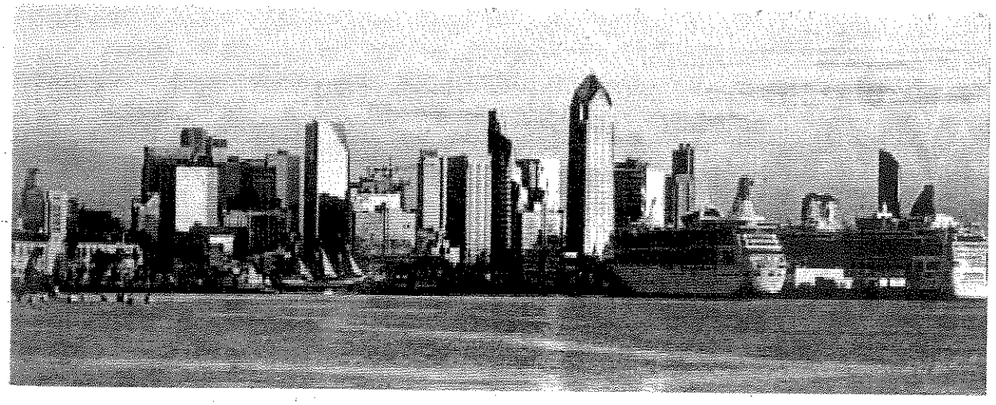
1. Basic clinical trial design for bioequivalence assessment;
2. Applications of active control trials with a placebo arm and a simple adaptive plan may be used to improve the study power of the 3-arm trial;
3. Hypotheses testing based decision rule and the confidence interval based decision rule for equivalence or non-inferiority testing in terms of difference between two proportions;
4. Equivalence testing or non-inferiority testing of trials with paired data;
5. Issues on data transformation;
6. Non-inferiority testing and general linear model.

About the Instructor:

Dr. Sue-Jane Wang is currently a senior mathematical statistician and serves in the Active Control Working Group and the CASE (Committee for Advanced Scientific Education) committee in FDA. She received an "Excellent in Communication" award from CDER, FDA. Dr. Wang received her M.S. in Biostatistics from University of California, Los Angeles and Ph.D. in Biostatistics from University of Southern California in 1993.

Dr. Yi Tsong is currently a mathematical statistician of Quantitative Methods and Research Staff of CDER, FDA. He specialize in postmarketing quality assurance and risk assessment. He chaired the CDER Office of Biostatistics Committee on Research and Training. He received CDER and FDA level awards for contributions in postmarketing drug risk assessment, medication errors, quality control evaluation, drug compliance, in vitro bioequivalence and for advisory on CDER postmarketing risk assessment external contracts. Dr. Tsong is currently the Executive Director of ICSA. He received his Ph.D. in Mathematical Statistics from The Univ. of North Carolina-Chapel Hill in 1979.

Drs. Wang, Tsong and Dr. James H.M. Hung received a FDA Outstanding Service Award and a CDER Scientific Achievement Award in 2002. The two awards were received based on their research contributions generated from a CDER funded regulatory science research project on active control non-inferiority clinical trials, of which Dr. Wang was the Principle Investigator.



INTERNATIONAL CHINESE STATISTICAL ASSOCIATION
YEAR 2003 APPLIED STATISTICS SYMPOSIUM

Registration Form

JUNE 22-24, 2003

University of San Diego, 5998 Alcala Park, San Diego, CA, 92110-2492

Name: (English) _____ (Chinese, if any)_____

Affiliation: _____ Affiliation category: Industry ☐ Gov. ☐ Academic ☐

Mailing Address: _____

_____

Phone: (___)_____ Fax: (___)_____ E-Mail: _____

Speaker: Yes ☐ No ☐

1. Symposium Registration

Please check the appropriate box:

Membership Type	By May 15, 2003	Check (√)	After May 15, 2003	Check (√)
Regular Member	$140	☐	$160	☐
Regular Nonmember	$180	☐	$200	☐
Student Member	$50	☐	$70	☐
Student Nonmember	$70	☐	$90	☐

2. ICSA 2003 Membership

For ICSA member attendees, please renew your membership by checking the appropriate box below. **Nonmember attendees will receive one-year membership.** Please print Membership Application Form from http://www.icsa.org and mail it with this Registration Form.

Annual ICSA regular membership:	$40	☐	Annual ICSA student membership:	$20	☐
Lifetime ICSA permanent membership:	$400	☐			

3. Short Course Registration* (Sunday, June 22 from 9:00 AM to 5:00 PM)

	By April 30, 2003	Check (√)	After April 30, 2003	Check (√)
Non-student	$300	☐	$350	☐
Student	$50	☐	$60	☐

*The short course registration fee includes breakfast, lunch, and coffee breaks.

Please select the short course you would like to attend:

Topic	Instructor	Check (√)
1. Practical Guidance of Generalized Linear Mixed Models	Professor Charles E. McCulloch, *University of California, San Francisco*	☐
2. Tutorial on Statistical Bioinformatics	Professor Jun Liu, *Harvard University*	☐
3. Bootstrap Methods; A guide for Practitioners	Dr. Michael R. Chernick, *Novo Nordisk Pharmaceuticals*	☐
4. Cancer Trials for Practitioners – Experimental Design & Efficacy Analysis.	Dr. Kao-Tai Tsai, *Aventis Pharmaceuticals*	☐
5. Active Controlled Clinical Trials	Drs. Yi Tsong and Sue-Jane Wang, *FDA*	☐

Note: A short course may be canceled due to lack of participation. This fee will be fully refunded in such a case.

4. Housing Request Form

Name of Attendee: _____

Arrival Date: _____

Departure Date:_____

If sharing room with another attendee, please give the name of other attendee_____

Room is guaranteed if the housing request form is received **on or before June 1, 2003**

Housing Description:

Rooms are in the Alcala Vista Apartments, University of San Diego (USD) (http://www.sandiego.edu/ugadmiss/campustour/tour32.html). Each apartment has two bedrooms, a living room and a kitchen. Each room has two single beds and a private bathroom. It is recommended to check in between 3:00-5:00 pm even though USD will accommodate late arrivals. Check-out is no later than 9:00am without prior arrangement. Penalty for Late check-out is one day room and board fee, for a lost key is $65.00, and for a lost meal ticket is $10.00 If you have any questions regarding the room rates, please contact Nancy Lo at Nancy.Lo@noaa.gov or 858-546-7123.

Housing Selection:

	Room Rates			Apartment Rates *(for families only)*			
	Single Occupancy	Double Occupancy		Apartment			
No. of attendees (meals included)	**1**	**1**	**2**	**1**	**1**	**2**	**2**
No. of guests (no meals*)	**0**	**1**	**0**	**2**	**3**	**1**	**2**
Total no. of people	**1**	**2**	**2**	**3**	**4**	**3**	**4**
1 Night	$88 ☐	$112 ☐	$144 ☐	$162 ☐	$192 ☐	$194 ☐	$224☐
2 Nights	$176 ☐	$224 ☐	$288 ☐	$324 ☐	$384 ☐	$388 ☐	$448☐
3 Nights	$264 ☐	$336 ☐	$432 ☐	$486 ☐	$576 ☐	$582 ☐	$672☐
4 Nights	$352 ☐	$448 ☐	$576 ☐	$648 ☐	$768 ☐	$776 ☐	$896☐

*Individual meal tickets for guests (non-attendees) can be purchased at the registration desk upon arrival.

Housing Cost (from table above) = $_____

For a group larger than 4, the charge for each extra guest is

$50/person single occupancy x ____people x _____ nights

$40/person double occupancy x _____people x _____ nights = $_____

Children under 5 years are free of charge

Total for housing cost: $_____

5. One-Day Tour To Tijuana, Mexico

A great opportunity for fun, shopping and sightseeing!! We have arranged for an exciting one-day tour to Tijuana, Mexico from 8:30 AM - 6:00 PM on Sunday, June 22. For detailed information see http://www.sandiegoscenictours.com/tjtours.htm or call 858-273-8687. A private group tour will be provided for reservations of a minimum of 25 people before May 15. Otherwise, you will tour as individuals. Price includes bus fare and tour guidance but not meals. Please submit your registration form **before May 15th**. Tourists are required to carry proper IDs: Photo ID's required for U.S. and Canadian citizens. Passport, I-94 Form, Multiple Entry Visa OR Resident Alien Card required for non-U.S. Citizens.

$28/adult x _____ adults = $_____
$15/child x _____ children (3-11) = $_____

Total Tour cost: = $_____

6. Total Conference Payment

Symposium Registration: = $_____
ICSA Membership: = $_____
Short Course: = $_____
Banquet on Monday night, June 23:
$30/person x ____ people = $_____
$10/child x ____ children (under 10 years old) = $_____
Total for Banquet = $_____
Housing: (from Housing Request Form) = $_____
Tijuana Tour: = $_____
Donation to ICSA: _____ = $_____

Total Payment: $_____
(Please make check payable to 2003 ICSA Symposium)

PLEASE SEND COMPLETED REGISTRATION FORM WITH CHECK TO:

Kathy Chi-Burris
Treasurer, ICSA 2003 Applied Statistics Symposium
c/o Pfizer Global Research & Development La Jolla Laboratories
11085 Torreyana Road
San Diego, CA 92121
Phone: (858) 622-7375
Fax: (858) 678-8248
E-Mail: kathy.chi-burris@pfizer.com

Cancellation Policy: Unless approved by the Committee, all symposium participants must register (except for keynote, dinner, and plenary speakers). Full refund for cancellation will be made if requested on or before May 15; 80% refund will be made if requested after May 15 but on or before June 1.

ICSA Student Awards & Travel Fellowships

The 13th Annual ICSA Applied Statistics Symposium will be held on June 22-24, 2003 at the University of San Diego in San Diego, CA. The Program Committee will again sponsor the Student Awards and Travel Fellowships. The main purpose of the award is to encourage student members of ICSA to participate and present their research work at this annual meeting.

Qualifications: The student must be an ICSA member (or join at the time of manuscript submission), a degree candidate in any term during 2003 at an accredited institute and be able to register and present the work at the 2003 symposium.

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

Review and Selection Process: Three review members of the Award Committee, appointed by the Chair of the Committee, will receive blinded copies of the submitted manuscripts from the Committee Chair and review them based on the following criteria: The manuscript should be well motivated by an application relevant to the specific field(s). The methodology developed should be applicable to the motivating problem.

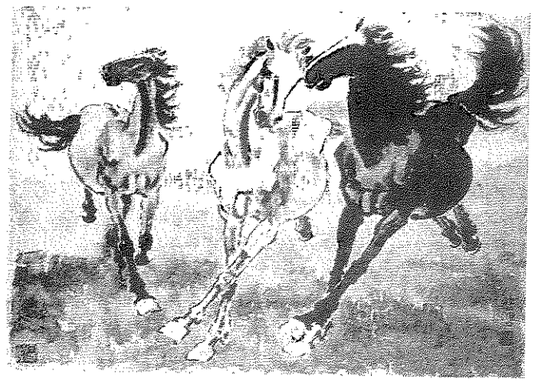
Inclusion of an application of the proposed methodology to a particular study will be favorably considered. Clarity of presentation in writing will be considered as well.

Up to 3 travel award winners will be selected by the Awards Committee chaired by Prof. Kung-Jong Lui. All winners will each receive a certificate, $400, and tuition for one short course of their choice. The winners will be notified by April 15, 2003.

Submission of Manuscripts: Manuscripts should be received and postmarked no later than February 28, 2003. The submission should include:
(1) A cover letter, (2) One complete title page with author(s), institutional affiliation, mailing address, phone/fax numbers and e-mail address, (3) Five copies of the manuscripts with only a title, but no information on authors or affiliation, on the first page, (4) Two copies of abstract, (5) Two copies of the ICSA membership application for non-members

Membership form can be downloaded from http://www.icsa.org

All materials should be mailed to:
Professor Kung-Jong Lui
(kjl@rohan.sdsu.edu)
Department of Mathematics and Statistics
San Diego State University
San Diego, CA 92182-7720



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 Yi Tsong, Ph.D. 13215 Lazy Glen Lane, Herndon, VA 20171			

Revised 8-29-2002

Georgia Southern University
Department of Health and Kinesiology
College of Health and Professional Studies

Position Tenure track Assistant Professor position in Biostatistics available August 1, 2003.

Responsibilities Coordinate and teach graduate courses in the Biostatistics emphasis of the Master of Public Health degree program. Provide academic and research advisement to graduate students. Conduct research in the area of Biostatistics and work in the newly established Center for Biostatistics. Secure external grants and contracts to support research and training programs. Promote and sustain interdisciplinary dialogue, research, and service. Hold active membership in a variety of professional associations/organizations. Serve on committees and participate in official activities at the departmental, college and university levels.

Qualifications An earned doctorate in Biostatistics or related field is required prior to the appointment date. Candidates must also have demonstrated potential for excellence in teaching and developing a research program in the area of expertise, as well as the ability to contribute to the University and profession through service activities. Expertise in SAS, clinical trials methodology and analysis, consulting in the pharmaceutical industry, experience working with a diverse student body, incorporating technology into instruction, and 1-2 years of college or university teaching experience in Biostatistics are preferred.

Salary Competitive and commensurate with experience.

Application Process: Send a letter of application addressing the qualifications cited above, curriculum vitae, and the names, addresses, and telephone numbers of at least five professional references to Dr. Jim McMillan, Chair Biostatistics Search Committee, Department of Health and Kinesiology, P.O. Box 8076, Georgia Southern University, Statesboro, GA, 30460-8076. Phone: 912-681-0200 email: jmcmillan@gasou.edu More information about the institution is available through <http://www.gasou.edu> and <http://chronicle.com/jobs/profiles/911.htm>. Screening will begin November 18, 2002.

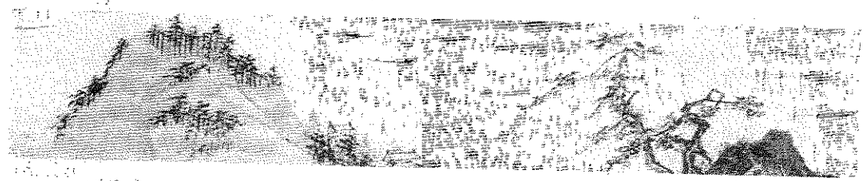
The names of applicants and nominees, vitae, and other non-evaluative information may be subject to public inspection under the Georgia Open Records Act. Georgia Southern University is an Equal Opportunity, Affirmative Action Institution. Individuals who need reasonable accommodations under the Americans with Disabilities Act in order to participate in the search process should contact the search chair.

International Chinese Statistical Association
Profit & Loss
January through December 2002

Ordinary Income/Expense		
Income		
Advertisement		200.00
Unrestricted Contributions Income		300.00
Membership Dues		16,140.00
Miscellaneous Income		5,841.93
Total Income		**22,481.93**
Expense		
Board Meeting		517.74
Casual Labor		46.50
Computer Hardware/Software		313.49
Contributions to ASA		500.00
ICSA at ASA meeting		
Banquet		271.22
Internet Registration		35.00
Licenses and Permits		110.00
Membership due refund		440.00
Supplies		
Office		773.23
Other		1,170.00
Postage and Delivery		
Ballot		725.05
Book/Journal Donation		91.00
Bulletin		6,666.51
Other		1,196.77
Total Postage and Delivery		8,679.33
Printing and Reproduction		
Jan. Bulletin		4,602.50
July Bulletin		3,500.00
Total Printing and Reproduction		8,102.50
Professional Fees for Tax filing		431.00
Program Expense		377.39
Web Page Hosting		1,200.00
Total Expense		**22,968.20**
Net Ordinary Income/Expense		**-486.27**
Other Income/Expense		
Interest Income		1,296.68
Net Other Income/Expense		**1,296.68**
Net Income		**810.41**

International Chinese Statistical Association
Balance Sheet
As of December 31, 2002

Assets		
Checking		9,094.75
Savings-CD		31,583.32
Savings-Money Market		33,887.05
Total Assets		**74,565.12**
Liabilities & Equity		
Equity		
Opening Balance 1/1/02		73,754.71
Net Income		810.41
Total Equity		74,565.12
Total Liabilities & Equity		**74,565.12**



International Chinese Statistical Association
Applied Statistics Symposium, 2002
Income & Expenses

Prepared by Yusong Chen

Income/Expense		
Income		
Symposium:		
Registration		24210.00
Banquet		3860.00
Donation		20095.00
Bank Credit		12.00
Membership		3680.00
Symposium Advances		3000.00
Short Courses:		
Registration		24400.00
Text Books		1320.00
Total Income		**80577.00**
Expense		
Symposium:		
Hotel		15085.80
Banquet		5590.00
Souvenir		4657.81
Student awards		1200.00
Advertisements		529.72
Membership-ICSA		4080.00
Bank Fee		74.30
Refund		712.00
General expenses		2027.70
Symposium advances refund		3000.00
Board Meeting		240.00
Short Courses:		
Instructors honorarium		4500.00
Hotel		6357.00
Equipment		432.00
Instructors travel expenses		1201.16
Text books		1545.99
Total Expense		**51233.48**
Net Income		**29343.52**

Note: Half of the short course net income to ICSA Treasurer ($5841.94) and the remaining balance to the Symposium General Treasurer ($23501.58).

Calendar of Meetings

June 22-25, 2003 WNAR Annual Meeting /IMS Western Regional Meeting
Location: Golden, Colorado, USA
Sessions include panels on general topics such as funding opportunities and the current shortage of biostatisticians, as well as technical sessions. Information available at: http://biosweb.njc.org/vc/biom /WNAR2003/WelcomePage.cfm

July 10-12, 2003 Bernoulli Society East Asian and Pacific Regional Conference
Location: Hong Kong.

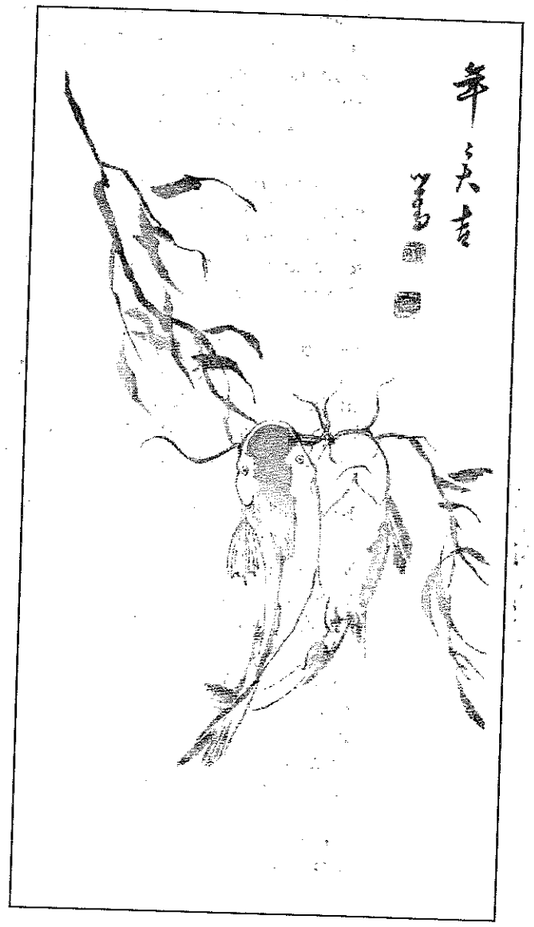
The Bernoulli Society East Asian and Pacific Regional (EAPR) Conference 2003 will be held at The Hong Kong University of Science and Technology (HKUST). The conference is organized by HKUST under the auspices of the East Asian and Pacific Regional Committee of the Bernoulli Society. Keynote speakers are David Aldous, Friedrich Gotze, Zhi-ming Ma, Wing Hung Wong and C. F. Jeff Wu.

Authors are now invited to submit contributed papers through email to eapr2003@ust.hk in either Latex or Microsoft Word formats. The deadline of the submission is 31 March 2003. For more conference information, please visit *www.bm.ust.hk /~eapr2003.*

July 24-26, 2003 - First Joint ISBA-IMS Meeting
Location: Intercontinental Hotel, Isla Verde, San Juan, Puerto Rico. More information at: http://www.cnnet.clu.edu/math/IMS-ISBA-PR2003/

July 29 - August2, 2003 – IMS New Researchers Conference
Location: University of California, Davis. Information available at: http://www-rohan.sdsu.edu /~ralevine/NRC/

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



Our Sincere Thanks!

The Editorial Team

This is a wonderful adventure for us to edit the ICSA Bulletin, as this makes us appreciate so much more of the wonderful friendship around the world. Many good friends have squeezed time from their busy schedule to write for the Bulletin under short notice, especially for T.Y. Lee, Jun Liu, Bill Wei, Zhiliang Ying, Yi Tsong, J.P. Liu, S.J. Wang, James Hung, and Ker-Chau Li.

We would like to congratulate G. Li and G. Zhi as well as friends from Singapore for being the pioneers in reporting the statistical activities in Mainland China and Singapore regions. Of course, the continual effort of Hailiang Yang to inform us the activities in Hong Kong is also highly appreciated. This truly makes our organization international. We hope to have an increase of activities around the world to report to you in the next issue.

We are also grateful to have the continued support of S.J. Wang, the past editor. Like a good mother, she continuously keeps her watchful eyes on the Bulletin even though it has grown older and left her nest. Please don't forget to send your contributed articles to her. She is taking care of the column of Controversial Statistical Issues.

We would also like to thank Bob Lippman for translating the writing of Zhung Ban-Chow included inside the back cover. The world would be a much better place if everyone were to remember and follow Zhung's philosophy.

But, as everyone knows, no chef can create any wonderful feast without the ingredients! We NEED everyone's input! We need everyone to send in their articles. We need everyone to send in their graphics of classical Chinese art piece such as painting, calligraphy, sculpture, antiques, and landscape photos. The quality of our Bulletin can only be raised with everyone's support. With the ingredients from everyone, we can definitely make the ICSA Bulletin a more enjoyable journal to read!

As we have said in the beginning, ***THIS IS OUR BULLETIN. PLEASE JOIN THE CLUB!***

It's hard to be wise,
Not easy to be a fool;
And for someone wise to behave like a fool is
very difficult.

Still, in life, one can ease up a bit,
One can back off a little.

If one has peace of mind,
There's no need to worry about the future.

(By Chinese Scholar: Zhung Ban-Chow)
鄭板橋