泛華統計協會

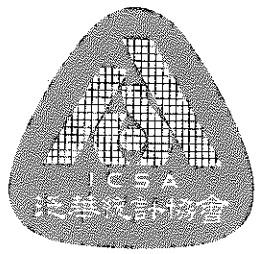
會刊

International Chinese Statistical Association

Website: http://icsa.org

Bulletin

January 2002



編者的話：

會刊是我們會員的主要交流園地，預計每年在一月及七月發行兩期，希望各位踴躍賜稿，文章以一至二頁爲主，須用中文繁體字或英文書寫，文章性質以一般性非專門學術的文章爲主，以前或本期會刊內的各篇短文都是很好的範例，稿件如採用，我們希望能由作者安排打字，完稿的上，下，左，右邊（margin）至少要有四分之三英吋，頁碼離底邊二分之一英吋，打字打在 8" x 11" 的白紙上。爲統一起見，若以英文書寫，請用 Microsoft Word 打字，中文繁體字請用 Microsoft Word with Valupack Fareast feature 打字。稿件的截止日期是六月十五日及十二月十五日。截止日期後收到的將列爲下一期的稿件。

刊內付有本會的申請表（application form），其中會員動態內的著作（publication）以一年內的新作品爲限。

從 1994 年度起本會酌收徵才廣告費一頁 U.S.$200 或半頁 U.S. $120。

對於以上的一些構想，如果您有任何建議，請與泛華協會執行長莊易（地址見下）聯絡，以期這份刊物更爲完善。

Website: http://www.icsa.org

I.C.S.A. c/o Yi Tsong, Ph.D.
13215 lazy Glen Lane
Herndon, VA 22071
U.S.A.

EXECUTIVES AND MEMBERS OF THE COMMITTEES OF ICSA 2002

EXECUTIVES

President:	William W. S. Wei (2002)
Past President:	Chao Agnes Hsiung (2002)
President-elect:	Zhiliang Ying (2002)
Executive director:	Yi Tsong (2001-03)
Treasurer:	H. M. James Hung (2001-03)

BOARD OF DIRECTORS

Ngai Hang Chan (2001-03), Chen-Hsin Chen (2001-03), Rong Chen (2002-04), Jianqing Fan (2000-02), Agnes Hsiung (2000-02), Mei-Ling Lee (2001-03), Guo-Ying Li (2001-03), Ker-Chau Li (2000-02), Zhaohai Li (2002-04), Kar K. Lin (2000-02), Jun Shao (2000-02), X. Don Sun (2001-03), Naitee Ting (2002-04), Mei-Cheng Wang (2000-02), Sue-Jane Wang (2002-04), William W.S. Wei (2001-03), Zhiliang Ying (2002-04), Heping Zhang (2002-04), Frank Shen (2000-02, Biometrics Section Representative).

STANDING COMMITTEES

PROGRAM COMMITTEE:
Wei-Yann Tsai (Chair 2002, member 2002-2003), Hubert J. Chen (2002), Zhiliang Ying (2002)
Term of reference: to recommend conference and symposium sites, including candidates for their chairs; to recommend general policy for all meetings, subject to approval by the Board of Directors.

FINANCE COMMITTEE:
H.M. James Hung (chair 2001-03), Xiu Chen (2001-03), Wei-Ying Yuan (2001-03)
Term of reference: to oversee the budget and to recommend long-term financial planning, including investments of the Association's assets, subject to approval by the Board of Directors. Term of reference: to oversee the budget and financial situation of the Association.

NOMINATING AND ELECTION COMMITTEE:
Naitee Ting (chair 2002, member 2002-03), Jeff C. F. Wu (2002-03), Jen-Pei Liu (2001-2002), Heping Zhang (2002) Term of reference: to nominate the candidates for President-elect and members of the Board of Directors.

PUBLICATION COMMITTEE:
James J. Chen (Chair 2002, member 2000-02), I-Shou Chang (2001-03), Jun Shao (2002-04), Sue-Jane Wang (Bulletin), Ker-Chau Li (Statistica Sinica), Yi Tsong (ex-officio)
Term of reference: to oversee the publication policy of the Association and make recommendations to the Board of Directors

CONSTITUTION COMMITTEE:
Frank Shen (Chair), Shein-Chung Chow, Chien-Pai Han Term of reference: to review the Association's Constitution and By-Laws and prepare a revision if necessary.

CURRENT COMMITTEES

MEMBERSHIP COMMITTEE:
Tzu-Cheg Kao (Chair 2002, member 2000-02), Rongdean Chen (2001-03), Chong Gu(2000-02), Zhaohai Li (2000-02), Xufeng Niu (2000-02), Ming Tan (2001-03),

Heping Zhang (2001-03), Ling Chen (2002-04), Wai-sum Chan (2002-04, Hong Kong), Chen-Hsin Chen (2002-04, Taiwan), Guo-Ying Li (2002-04, China)
Term of reference: to recruit new members and contact interested potential individuals and organizations.

FUNDRAISING COMMITTEE:
Alice Hsuan (Chair 2002, member, 2001-03), Jianping Dong (2002-02), Kuang-Chao Chang (2002-02)
Term of reference: to consider fundraising drives through individuals and corporations

PUBLIC RELATIONS COMMITTEE:
Yi Tsong (Chair, 2002, member 200-02), Naisyin Wang (2000-02), Shi-Yong Feng (China), Sik-Yum Lee (Hong Kong), Lung-An Li (Taiwan)
Term of reference: to contact the news media and publicize ICSA activities; to serve as a liaison between ICSA and other professional organizations such as ASA, Biometric Society for joint activities.

AWARDS COMMITTEE:
Lynn Kuo (Chair 2002, member 2002-02), Wen-Jang Huang (2001-03), Zhaohai Li (2001-03), Jane-Ling Wang (2002-04), Ming Tan (2000-02), Sue-Jane Wang (2002-04)
Term of reference: to accept, evaluate, and recommend nominations for ICSA various awards.

COMMUNICATION COMMITTEE:
Rong Chen (Chair 2002, member 2002-04), Don Sun (Web), Hubert Chen (Listserv)
Term of reference: to evaluate the database and use of internet.

APPLIED STATISTICS SYMPOSIUM COMMITTEE:
William W. S. Wei (Chair), Danny Chaing, Ivan Chan, George Chao, Yusong Chen, Alica Hsuan, Lee Huang, Frank Shen
Term of reference: to organize the Applied Statistics Symposium, 2002.

BOOK AND JOURNAL DONATION COMMITTEE:
Tar Timothy Chen (Chair)
Term of reference: to solicit book and journal donations and to arrange their delivery to universities or colleges in need.

SYMPOSIUM PLANNING COMMITTEE:
WeiChung J. Shih (Chair 2002, 2001-02), H. M. James Hung (2002-03), Rong Chen (2001-03), Tar Timothy Chen (2001-02), Jiann-Ping Hsu (2001-03), Zhilian Ying (2001-03)
Term of reference: to recommend future symposium site to the Board.

STRATEGIC COMMITTEE: (all former presidents)
Chao Agnes Hsiung (Chair 2002), Chien-Pai Han, Tar Timothy Chen, Jeff C. F. Wu, Shein-Chung Chow, Kuang-Fu Cheng, Smiley Cheng, Chiao Yeh, Yuan A, Chow, Jack C. Lee, Grace Yang, Jia-Yeong Tsay, James Fu, Georhe Tiao
Term of reference: to plan long-term strategies for the Association

BIOMETRICS SECTION (2001)

H. M. James Hung (Chair), Wei-Chung J. Shih (past chair), Jen-Pei Liu (Chair-elect), Shou-en Lu (secretary), Gang Li (treasurer), Frank Shen (ICSA Representative 2000-02)

EXPRESS YOUR OPINION .....

Dear ICSA Members,

We added a list of some upcoming statistical meetings. In addition, the upcoming year 2002 applied statistics symposium and year 2003 applied statistics symposium are sponsored by the ICSA. We'd like to notify you ahead of time and encourage your involvement by contacting the chairs of the corresponding events (details see announcement) or by sharing your research findings with the members. Your suggestions / comments are welcomed. Details inside ....

IN THIS ISSUE ...

GET INVOLVED IN NEXT ISSUE .....

Special topic – Weather
Controversial statistical
Issue – Bridging Studies

Get involved in the next issue by email your articles to the Editorial Board at
WANGS@CDER.FDA.GOV

Editor's Page

Contributions and Highlights

In this issue, we have a guest editorial served as a commentary to the controversial statistical issue – missing data. In addition, we introduce pharmacogenomics as the special topic of the issue. A few highlights of the Bulletin follows. Professor George Tiao, a founding member of ICSA, has been nominated to be a candidate for the 2003 ASA Presidential race, a special interview article prepared by many of our thoughtful long time members and written by Professor Ruey Tsay is included. William Wei, our President, has asked the constitution committee to take a serious read of the constitution and the by-laws, a revised version including the before and the after text is published. As usual, several members contributed to the Statistics Delights.

Sue-Jane Wang
Editor-in-Chief

What is Missing in the Missing Data Controversy

This guest editorial, serving as a Commentary of the controversial statistical issue – missing data, is by Office Director of Biostatistics, Center for Drug Evaluation and Research, US Food and Drug Administration, ***Robert T. O'Neill****, who is a lifetime member of ICSA. Dr. O'Neill has been leading the biostatistics group within the FDA in the new drug approval process of clinical trials for more than 20 years?? and has a long standing interest in the missing data implication of the clinical trial planning, its conduct, data analysis and final interpretation.*

Clinical trials are the primary experimental mechanism used to establish the efficacy and safety of a new medical product or to compare the relative benefits of two or more products. Two critical components of the infrastructure of the modern clinical trials are the prospective statistical planning of the study design and the prospective planning for the subsequent implementation of a variety of statistical data analyses that produce the quantitative estimates and uncertainties associated with estimates for the
(See the column of Controversial Statistical Issue)

MESSAGE FROM THE PRESIDENT

Dear ICSA Members: January 2002

ICSA was established in 1987. For an organization, fifteen years is a relatively short time. However, through the devotion of its many members and officers, the ICSA has grown steadily and become very well recognized and respected in the statistical profession.

Although most of our ICSA members are in the US (63%), we also have a significant number of members in Asian regions such as Taiwan (16%), China (10%), Hong Kong (4%) and Singapore (2.5%). ICSA is clearly a US based international organization. Thus, ICSA promotes statistics education, research, and applications not only in the US but also worldwide particularly in Asia to provide service to our members. This is the exact reason that we organize conferences in both the US and Asia. The yearly Applied Statistics Symposium began in 1990 on the East Coast and moved to Chicago in 2001. Although it will be held again this year on the East Coast, it will move west to San Diego in 2003. To serve our members in Asia, we have also organized international conferences. The tri-year ICSA International Conference was held in 1990 (Hong Kong), 1993 (Taiwan), 1995 (Beijing), 1998 (Kun Ming), and 2001 (Hong Kong). In addition, ICSA also publishes the highly regarded journal, *Statistica Sinica*, jointly with the Institute of Statistical Science of Academia Sinica in Taipei, to serve the statistical community.

To further enhance our service, I would like to share the following thoughts with our members:

- As an international statistical organization, we should keep our members not only informed but more importantly actively interested in worldwide statistical activities especially in regions such as China, Hong Kong, Singapore, and Taiwan where there are a significant number of ICSA members. This is one of the reasons that the *ICSA Bulletin* has a special column for Regional Activities. It is hoped that with the help of volunteers, we can publish more regionally featured reports and articles for our members to share their interests and encourage interaction among these areas. The experiences that Professor Tiao shares with us in the interview report in the current issue should also motivate our members to help promote statistics in these Asian regions.
- Building upon the success of the ICSA International Conferences, we can further promote the field in Asian regions by forming regular joint statistical meetings with the region's statistical organizations. Through these corporations, we can more effectively promote the use of statistics in business, industry, and government in these regions.
- Since the US has very highly advanced research and applications in statistics which may be of interest to our international members, ICSA can arrange some useful study tours for our colleagues from China, Hong Kong, Singapore, and Taiwan, especially during the ICSA yearly Applied Statistics Symposium or Joint Statistical Meetings. Through these tours to educational institutions, government agencies, and industrial research centers, they can observe American methods of educating statisticians, developing statistics research and solving statistical problems, increasing productivity, and helping decision making.
- There are many statistical societies, and most statisticians join more than one. We encourage our members to participate actively in the activities of the societies to which they belong. One of our founding members, Professor Tiao, is a candidate for the 2003 ASA president. We wish him the best of luck.

I wish you a Happy New Year,

William Wei
President

Special Thanks From the Editorial Board

The Editorial Board would like to thank Drs. Timothy Chen, William Wei, Sue-Jane Wang, Kung-Yee Liang, Yi Tsong, Grace Young, etc. for their enthusiastic effort during the preparation of the special interview.

If you have a new idea and are interested in joining us, please send your C.V. including your plan to the Editorial Board WANGS@CDER.FDA.GOV for consideration.

FROM THE EXECUTIVE DIRECTOR

Dear friends:

Happy New Year! ICSA is in its 15th year now! With the effort of our Board of Directors, Committees and support from all of you, our organization enjoyed a steady growth throughout these years. Year 2001 has been a very successful year under the steeling of President Agnes Hsiung. We also regret to see the departure of our Past President Chien-Pai Han, Board Directors, I-Shou Chang, Rongdean Chen, Dan-Yu Lin, Nancy C. H. Lo, Zhiliang Ying. I want to take the opportunity to congratulate them for completing a very successful term. I am also regretting to see their leave from the positions. But, old soldiers will never die; they will come back to serve at different positions. We also need to congratulate Tar Timothy Chen for running an extremely successful and efficient journal donation program. Although retired from his teaching position, he will continue to chair this program.

I am happy to announce that we have a very successful election of the executives and Directors in the summer. It was a very suspenseful election and every candidate received balance ballot till the end. However, we don't have the Florida controversial this year. I am very glad to inform you that our newly elected executives and Directors are

Ziliang Ying (president-elect 2003)
Directors: Naitee Ting, Sue Jane Wang, Rong Chen, Heping Zhang, Zhaohai Li
Jen-Pei Liu (Biometrics Section Chair)

The result does prove my statement about the old soldiers. Welcome on Board!

Recently I mailed the 2002 membership renewal and directory information update form. If you renewed your membership before June 1 last year, your ICSA membership are likely overdue. Please make sure to renew your membership timely. When you update your directory information, please make sure that you include the most updated e-mail address. For information changes, you may e-mail the form to me.

ICSA member wants to subscribe the following journals may enjoy the following privileged rates.

Statistical Methods in Medical Research $112
Statistical Modeling Privileged rate $100.

To find out more details and their complete subscription rates please visit journals website at www.arnoldpublishers.com/journals. This arrangement was made by Mr. Patrick Kelly, Editors of the journals.

Please take you time to visit your association's web site and many timely information will be posted there. Hope you have a Great New Year!

Sincerely yours,
Yi Tsong,
301-827-3206, Fax (301) 480-2825,
email TSONG@CDER.FDA.GOV

Minutes of 2001 ICSA Membership Meeting

At Grand Buffet, Atlanta, Georgia on August 5th, 2001

Time: 7:00 -7:30 PM
Chair: William Wei (President Elect)
Minutes: Yi Tsong
Attendees: About 70 ICSA Members

1. ICSA Award Presentation
   William Wei presented the following awards at the meeting:

   Professor Chien-Pai Han was presented with the 2001 ICSA Distinguished Service Award with the following citation:

   > Service on the very first Board of Directors and on various committees, significant contributions and effective leadership as the President, teaching, research and editorial activities, promotion of closer relationships between ASA, COPSS, and ICSA, and long-range planning for ICSA symposiums and conferences.

   Dr. Irving K. Hwang was presented with the 2001 ICSA Distinguished Service Award with the following citation:

   > Service on the Board of Directors, instrumental contributions in creating and maintaining the Biometrics Section and Applied Statistics Symposiums, significant contributions in fund raising activities, significant contributions in the pharmaceutical industry, and significant contributions in helping members in their career development.

   Dr. Naitee Ting was presented with the 2001 ICSA Distinguished Service Award with the following citation:

   > Multiple significant contributions while serving as the Executive Director, significant contributions in the publication of ICSA Bulletins, and significant contributions in the annual elections.

2. Report from the president (by William Wei)

   William Wei indicated that President Agnes Hsiung was preparing for her attendance of the 5th ICSA International Conference and expressed her regret that she couldn't attend the meeting. He was asked to present this report. William first thanked the Board of Directors, Committee members and Editors of ICSA Bulletin and Statistica Sinica for their important contributions in 2001. He congratulated the organization committees of the Applied Statistics Symposium and the 5th ICSA International Conference for successful and well-received programs. He pointed out that the 2001 Applied Statistics Symposium (co-chaired by Prof. Rong Chen and Dr. Jiann-Ping Hsu) held in Chicago was a great success. The

symposium attracted seventy some non-member participants (including new members and membership renewals). The 2002 Applied Statistics Symposium will be held in greater Philadelphia between June 6 and 8, 2002. The 5th ICSA International Conference will be held from August 17 to 19 in Hong Kong.

The President-Elect said that 2001 was another successful year for ICSA as shown in our successful conferences, journal and bulletin. In addition, many members received great honors from professional societies. He pointed out especially the following honorees:

- Prof. George Tiao received the ASA 2001 Wilks Memory Award and the 2001 Shiskin's Award
- Prof. Xiao-Li Meng received the 2001 COPSS Award
- Profs. Kai-Tai Fang, Jason Hsu, Kung-Jong Lui, Wei Yann Tsai, and Naisyin Wang were elected as new ASA Fellows
- Prof. Qi-Man Shao elected as a new IMS Fellow

Finally, he expressed sincere thanks to the local organization committee for the excellent arrangement of the banquet meeting. The committee was led by Profs. Hubert Chen, Yu-Sheng Hsu, Xiang-Rong Yin, and many other members from the University of Georgia and Georgia State University.

3. Business Report

Executive Director (Yi Tsong)

Yi Tsong expressed his inexperience in taking over a well-performed job from Naitee Ting this year and asked for patience and help. He indicated that the management direction would be online bound and looked forward to seeing the further progress of the state-of-the-art ICSA homepage. He also reported briefly for other executives and committee chairs. He reported the results of the election.

The newly elected executives and Directors are

President-elect (2003):	Ziliang Ying
Board Directors:	Naitee Ting Sue-Jane Wang Rong Chen Heping Zhang Zhaohai Li
Biometrics Section Chair:	Jen-Pei Liu

Meeting was adjourned at 7:30 for banquet.

THE REVISION OF THE ICSA CONSTITUTION AND BY-LAWS

It has been nine years since the ICSA Constitution and By-Laws were revised. To meet the Association's changing needs, I recommended Shein-Chung Chow, Chien-Pai Han, and Frank Shen (Chair) to 2001 President Agnes Hsiung in November 2001 for appointment to the Constitution Committee to review the Association's Constitution and By-Laws. After careful study, the Committee submitted several changes to the Board of Directors and received the approval of the Board to publish the proposed revision in the January 2002 issue of the Bulletin. We sincerely invite your comments on the proposed revision highlighted in blue as shown below. The Board will carefully consider your suggestions at the next Board meeting in June 2002. Thank you for your support.

William Wei

CONSTITUTION OF THE INTERNATIONAL CHINESE STATISTICAL ASSOCIATION

ARTICLE I. NAME

The name of this organization is the International Chinese Statistical Association, hereafter called the Association.

ARTICLE II. OBJECTIVES

The Association is a non-profit organization. The Association is organized, and will be operated, for educational, charitable, and scientific purposes only. Its objectives are:

* to promote the theory and applications of statistical disciplines through scholarly activities, including publication of journals in statistics and probability, scientific meetings, and other educational programs;

* to broaden applications of statistical techniques in all areas of society, including industry and government;

* to promote better understanding and interest by the general public in statistical methodology and related applications;

* to promote better communication through the development of standards and common terminology;

* to foster cooperative efforts among educational, research, industrial, and governmental personnel in statistical activities.

The objectives are pursued without regard to race, creed, color, sex or nationality.

ARTICLE III. MEMBERSHIP

The Association may have different categories of members, including individuals and organizations. An individual or organization who is interested in the objectives of the Association may apply to become a member. The right to vote, to sign referendum petitions, to hold office, and to sign nominating petitions shall be limited to individual members.

ARTICLE IV. SECTIONS AND DISTRICTS

Sections and Districts shall be encouraged to develop and explore ways to vigorously pursue the objectives of the Association, including meeting on matters of specialized interest such as current research and findings in a specific area, expository sessions on a single topic, or in-depth discussions of applications in a new field or of a new type.

(i) Sections: Sections may be established by the Board of Directors as provided in the By-Laws in order to promote the objectives of the Association. Each Section shall cover a field of statistical methods, theory, or applications which is sufficiently broad to represent active interests within the scope of the Association. Each member may belong to one or more Sections.

(ii) Districts: Geographic Districts shall be established by the Board of Directors as provided in the By-Laws. Each member shall belong to one and only one District based on the member's mailing address.

Only a member of the Association can hold membership in a Section or a District.

ARTICLE V. BOARD OF DIRECTORS

The Board of Directors is the policy-making and legislative body of the Association. Its actions are subject to the referendum of individual members as indicated in the By-Laws. The Board of Directors shall consist of at least nine (9) but no more than twenty (20) members, including the President, President-Elect, Past President, one member from each Section and each District; hereafter the members will be called the Directors.

ARTICLE VI. OFFICERS

The officers of the Association shall be the President, President-Elect, Past President, ~~Secretary~~ Executive Director and Treasurer. Executive Director ~~Secretary~~ and Treasurer are appointed by the Board of Directors and shall be eligible for immediate reappointment after completing a full term as described in Article VIII.

The officers of each Section shall be the Chair, Chair-Elect, and Secretary-Treasurer.

ARTICLE VII. METHOD OF ELECTION

Only individual members of the Association may become candidates for offices and Directors. All individual members shall be eligible to vote for the positions of President-Elect, President and Directors for the initial term or if such a nominations is made under provisions of the By-Laws. Only the individual members in each Section shall be eligible to vote for the officers and Director of that Section. Only individual members of each District shall be eligible to vote for the position of Director from that District.

ARTICLE VIII. TERMS OF OFFICE

Once elected to the position of President-Elect, the incumbent shall serve a three-year term. The first year the incumbent shall serve as President-Elect, the second year as President, and the third year as Past President. No Past President shall be eligible for immediate re-election to the office of President-Elect or President after the completion of the term.

Approximately one third of the members of the Board of Directors, excluding the President, President-Elect, and Past President shall be elected annually to serve a three-year term. In addition, each Section and each District shall elect one Director for a three-year term, except the Board of Directors may provide initial terms of one or two years in order to facilitate election of an approximately equal number of such Directors, and final terms of one or two years in order to facilitate any changes of Sections or Districts. No Directors completing a full term shall be eligible for immediate re-election to the same office.

Terms of office of five years or less shall be determined by the Board of Directors for the Executive Director ~~Secretary~~ and Treasurer. The Chair-Elect of each Section shall serve a one-year term, and at the end of this term automatically become Chair for a one-year term. No chair shall be eligible for immediate re-election to the office of Chair-Elect of the same Section.

Terms of office shall end, and new terms shall begin on January 1, but each office holder shall serve until a successor takes office. No individual may serve in two capacities on the Board of Directors.

ARTICLE IX. COMMITTEES

The Committees of the Association shall consist of the Standing Committees, which are named in the By-Laws, and such Current Committees as may be established or dissolved by the President with the consent of the Board of Directors as the situation warrants.

ARTICLE X. PUBLICATIONS

Publications of journals and other periodicals, reports, proceedings or other publications may be authorized in the By-Laws or by vote of the Board of Directors.

ARTICLE XI. MEETINGS

The Association shall have at least ~~a general~~ one membership meeting each year held at a time and place designated by the President with the consent of the Board of Directors.

ARTICLE XII. AMENDMENTS

Amendments to the constitution may be proposed by the Board of Directors or by a petition signed by ten percent (10%) of the Association's individual members. An amendment originated by petition shall be referred to the Board of Directors for its recommendation as to ratification. Regardless of the recommendation of the Board of Directors, an amendment proposed by petition must be submitted for vote at a membership ~~general~~ meeting. At least thirty (30) days' written notice shall be given, and a copy of the proposed amendments shall be sent with such written notice, to the membership of the meeting in which there is to be a vote to amend the Constitution. A two-thirds affirmative vote of the members voting shall be required for ratification, provided that the number of affirmative votes exceeds twenty percent (20%) of the Association's individual members.

BY-LAWS OF THE INTERNATIONAL CHINESE STATISTICAL ASSOCIATION

ARTICLE 1. MEMBERSHIP

1.1 Members. Members are classified into the following categories:

Individual members. An individual member shall be a person interested in the objectives of the Association whose application for membership is approved by the Board of Directors. The Board of Directors may delegate the function of approval.

Organizational Members. Institutions, corporations and other organizations interested in the objectives of the Association may be admitted to organizational membership by vote of the Board of Directors. The Board of ~~d~~Directors may delegate the function of approval.

1.2 Membership Year. The membership year for a member shall be the calendar year in which the member pays his/her membership dues.

1.3 Termination of Membership. Privileges of membership in the Association shall be automatically suspended if a member has failed to pay his/her dues for twelve months following the mailing of the first renewal notice. Membership may be terminated upon a finding by a two-thirds vote of all the members of the Board of Directors that a member has acted in a manner detrimental to the Association, provided that prior to voting the ~~b~~Board has given written notice to the member describing the charges against the member, and there has been due opportunity for the member or his/her designee to respond and a hearing by a committee appointed by the Board of Directors has been conducted.

ARTICLE 2. FINANCE

2.1 Dues and Subscriptions. The Board of Directors shall have the responsibility of determining the schedule of membership dues (both individual and organizational) and subscription rates. This schedule may provide for a special rate to students for a limited number of years, to members fully retired from employment, to husband and wife in case they agree to receive a single copy of publications and notices, and to such other groups as designated by the Board of Directors. Any new schedule of dues and rates shall be announced in a newsletter and shall become effective at the beginning of the next calendar year unless, within a period of four weeks after the mailing of a newsletter, a referendum is requested as provided in ARTICLE 4 of these By-Laws.

2.2 Fiscal Year. The fiscal year shall be the calendar year.

2.3 Financial Authority. All funds of the Association shall be deposited with the Treasurer, who shall make disbursements therefrom under regulations of the Board of Directors. With the approval of the Board of Directors, the Treasurer may delegate to an assistant the powers aforementioned as well as the power to sign checks, and access to safe-deposit boxes.

2.4 Publication of Financial Reports and Audit. The Treasurer shall submit to the Board of Directors, within one month of the end of each half fiscal year, a statement of the Association's current financial condition, including assets, liabilities, income and expenditures. This may be done at a meeting or by mail. The Treasurer shall also make a financial report to the Board of Directors within two months after the end of each fiscal year. This annual report shall be audited by a professional accountant selected by the Board of Directors. The auditor's report shall be published with the Treasurer's report in the next available newsletter.

2.5 Financial Relationship with Members or Units of the Association. The Association shall not be responsible for the debts or expenditures of any of its members or units (e.g., Sections, Districts, etc.) unless such debts or expenditures are authorized by the Board of Directors.

2.6 Financial Relationship with Cooperating Societies. The Board of Directors may delegate to the ~~Secretary~~ Executive Director or Treasurer the authority to negotiate financial arrangements with cooperating societies in connection with publications or other joint activities, subject to approval by the Board of Directors within the limitations provided in ARTICLE 5, Section 5.3 of the By-Laws.

ARTICLE 3. MAIL BALLOT

3.1 Quorum. In any mail vote of the Association's membership, all ballots received within a period set by the Board of Directors shall be counted and considered a quorum.

3.2 Balloting. If an election involves two candidates, the voter may cast only one vote for each position. If an election involves more than two candidates, the system known as approval voting will be used. Regardless of how many candidates there are or how many positions are to be filled, under approval voting the voter may vote for any number of candidates, but may not cast more than one vote for a candidate. Winning candidates are those with highest numbers of votes. Any tie will be broken by random selection conducted by the Board of Directors.

ARTICLE 4. REFERENDUM

Upon petition of at least 25 individual members of the Association, any action of the Board of Directors shall be subject to a referendum of the membership. The proposed referendum shall be published in a newsletter. Within 30 days after publication, a mail ballot of the individual members shall be taken. The will of the membership as expressed by a majority of those voting shall govern.

ARTICLE 5. ASSOCIATION WITH OTHER ORGANIZATIONS

5.1 Definition. A cooperating or associated society is a nonprofit organization interested in the objectives of the Association and concerned with the advancement of the statistical methodology or of its application.

5.2 Procedure. The Board of Directors of the Association may enter upon cooperative working arrangements with such organizations for promoting the objectives of the Association. Such cooperative working arrangements may include:

a. The exchange of representatives or delegates to each others governing bodies of working committees;

b. The assignment of Association representatives to an allied body composed of two or more societies including the Association;

c. Cooperative administrative, secretarial, financial, conference and publications functions, and other cooperative working arrangements.

5.3 Limitations on the Procedures. No cooperative working arrangement may be entered into by the Board of Directors which:

a. Results in the Association losing its identity as a separate organization;

b. Violates any article of the Association's constitution or By-Laws;

c. Requires the Association to allocate more than 5 percent of its annual revenue during any calendar year without receiving the equivalent in goods and/or services.

ARTICLE 6. OFFICES

6.1 Methods of Nomination. Three months before the annual meetings, the Nominating Committee shall:

a. Submit at least two nominations for President-Elect and, if the current President-Elect is vacated before February 1, two nominations for President.

b. Submit at least twice as many nominations as the number of members to be elected to the Board of Directors, the number of which shall be determined by the current Board of Directors under the provision of the Constitution, Article VIII.

c. Submit at least two nominations from each District for District Director, if the Director's three-year term is expiring at the year end.

The Nominating Committee is charged with full responsibility of the submission of nominations as described in the preceding paragraphs. Additional candidates on the ballot ~~nominations~~ for any elected office may be made by a petition signed by at least ten percent (10%) of the Association's individual members. ~~Twenty-five (25) individual members or by five (5) Directors.~~

Two months before the annual meeting, the ~~Secretary~~ Executive Director shall mail to individual members a ballot for the election to office from the persons nominated, along with a brief biographic sketch of each nominee.

The ~~Secretary~~ Executive Director, the Treasurer and the Editors of each periodical shall be appointed by the Board of Directors. If an initial procedure is required for a new office, the Board of Directors shall determine the procedure for nomination and election for initial terms.

6.2 Vacancies in Office. Except as provided for below, or in ARTICLE 6, Section 6.1, the Board of Directors shall fill any vacancy on the Board which occurs between elections.

If a vacancy occurs in the office of President, the President-Elect shall become President for the remainder of the current term as well as for the entirety of the succeeding term. If a vacancy occurs in the office of President and there is no President-Elect at the time, the Board of Directors shall choose a President from among the Directors to serve for the remainder of the current term and an election will be held for President for the following term.

If the office of President-Elect is vacated prior to February 1 and is not due to the President-Elect assuming the office of President, a new President and a new President-Elect shall be chosen as part of the annual election. Unless conditions occur as described in the preceding paragraph, if a vacancy occurs on or after February 1, and is not due to President-Elect assuming the office of president, the ~~b~~Board of Directors shall choose a President-Elect from among the Directors, who will then become President for the succeeding term.

If a vacancy occurs in the office of Past President, the office shall remain vacant for the remainder of the term.

6.3 Duties. The President is the chief officer of the Association and shall preside at all Association ~~general~~ membership meetings and at all meetings of the Board of Directors. Except as provided otherwise by the Constitution or the By-Laws, the President shall appoint the members of the committees of the Association and determine which member of each committee shall serve as its chair. All appointments shall be made prior to July 1. No appointment shall be made on or after July 1 without the consultation of the President-Elect.

The Past President shall serve as a special advisor to the President and a member of the Board of Directors. The President-Elect shall serve as a member of the Board of Directors and shall act as President in the event of the latter's absence or inability to serve.

The ~~Secretary~~ Executive Director and Treasurer shall be responsible for the duties assigned by the constitution and the By-Laws and for carrying out the policies determined by the Board of Directors. The ~~Secretary~~ Executive Director shall prepare an annual report for the publication in a newsletter on the activities of the Association as a whole.

ARTICLE 7. BOARD OF DIRECTORS

7.1 Members. Members of the Board of Directors shall be chosen as provided for in ARTICLE 6 of the By-Laws.

7.2 Meetings. The Board of Directors shall meet at least once a year. Meetings shall be held at the call of the President or on written petition signed by at least five members of the Board of Directors. ~~Secretary~~ Executive Director should mail the agenda of a meeting to the Directors at least two months prior to the meeting date.

7.3 Powers and Duties. Except as otherwise provided by the Constitution or the By-Laws, actions of the Board of Directors shall require a majority of those voting and the presence of a quorum. The quorum for the Board of Directors is a majority of its members. If a Director is unable to attend a meeting, he/she should appoint a current member of the Association to be his/her designee at the meeting, provided the name of the designee is sent to the ~~Secretary~~ Executive Director at least one month prior to the meeting date.

As the policy-making and legislative body of the Association, the Board of Directors shall make all decisions of policy. It shall adopt rules for the conduct of its business in harmony with the constitution and By-Laws; shall appoint representatives to cooperating societies and other agencies; shall schedule scientific meetings and other activities of the Association.

ARTICLE 8. SECTIONS, CHAPTERS, AND DISTRICTS

8.1 Sections. Sections may be established by the following procedure: Any group of 25 or more individual members of the Association (hereinafter called the sponsors) shall prepare a proposal for the formation of the new Section. This should include a charter and a documentation of purposes for the new Section.

The proposal and petition will be submitted to the Board of Directors for review and comment. The Board will then sent the proposal to each existing Section for comment.

Following review by the Board and receipt of comments from the existing Sections, the proposal will be returned to the sponsors for possible revision. The Board of Directors must consider the final proposal for approval or disapproval. If approval is not granted, the ~~b~~Board should provide guidance as to the reason for this action. If approval is given by the Board, at the time of the next annual election, the members will be asked to designate whether they desire to be members of such a new Section by singing a petition to that effect. (The petition will be prepared by the sponsors.) If at least 5 percent of individual members sign such a petition, a new Section shall be created as of the beginning of the next calendar year. The President, with the consent of the Board of Directors, shall appoint the officers of the Section to serve until officers are elected by the Section members at the time of the next annual election.

The Board of Directors may dissolve a Section if (a) it has become inactive, (b) the membership is less than 5 percent of the total membership of the Association, (c) at least 25

percent of its members request dissolution, or (d) it is determined by the Board of Directors that its continuance would be detrimental to the best interest of the Association.

8.2 Chapters. Each Chapter shall be governed by a constitution which is consistent with the Constitution and By-Laws of the association and provides for an annual meeting and annual election of officers. A copy of the Chapter constitution together with all subsequent amendments must be filed with the ~~Secretary~~ Executive Director of the Association.

All individual members residing in the area served by a Chapter shall be eligible for membership in that Chapter. Other persons may join the chapter as local associates. Only individual members of the Association may serve as President or Secretary of a Chapter.

8.3 Districts. Districts shall be established by the Board of Directors according to the geographic location and Directors must review the District boundaries at least once every 10 years. Taking into account whatever changes have occurred in the geographic distribution of the membership during this period, the Board of Directors may revise the districts when the Board deems it desirable to do so.

ARTICLE 9. COMMITTEES

9.1 Types of Committees and Terms of Memberships. The Committees of the Association shall consist of the Standing Committees as provided by the By-Laws and such Current Committees as may be established by the President with the consent of the Board of Directors. Current Committees may be established in order to carry through one project or to give voice to the Association in areas of special interest. Each committee shall prepare an annual report to the Board of Directors. Current Committees established in accordance with this article may be dissolved at any time by majority vote of the Board of Directors.

The terms of membership on Standing and Current Committees, when not specified in these By-Laws, shall normally be three years. Except for ex-officio members, no member may serve on a committee for more than six consecutive years without Board approval. Members of committees shall serve until their successors are appointed or elected.

Each Committee shall be governed by a Chair appointed by the President. All members of Standing and Current Committees shall be individual members of the Association.

9.2 Standing Committees. The Standing Committees are listed below.

A. Program Committee. The Program Committee for a given year shall consist of the Chairs of the Annual Meeting Committee, the Applied Statistics Symposium Committee and the International Conference Committee for the past two meetings. One of the members will be appointed yearly by the President as the Chair of the committee.

~~Previous year's Chair of the Committee and two individual members appointed by the President.~~ The Committee shall assist the Annual Meeting Committee for planning the annual meeting, recommend symposium and conference sites including candidates for their chairs and recommend general policy for all meetings, subject to approval by the Board of Directors. ~~Be responsible for planning, coordinating and arranging the annual meeting.~~

B. Finance Committee. The Finance Committee shall consist of the Treasurer as Chair, past treasurer, the Applied Statistics Symposium Treasurer, the International Conference Treasurer, and two individual members appointed by the President. The committee shall oversee the budget and recommend long-term financial planning including the investments of the Association's assets, subject to approval ~~Its duties shall be determined~~ by the Board of Directors.

C. Nominating and Election Committee. The Nominating and Election Committee shall consist of ~~four~~ six members appointed by the President with equal representation among academia, industry, government when possible and with the consent of the Board of Directors. The Committee shall make nominations for office as provided in ARTICLE 6 of the By-Laws and certify to the Board of Directors the outcome of the voting upon candidates for office and upon referendum to the members. As needed, the Committee shall also concern itself with developing mechanisms to insure proper conduct of elections. The ~~four~~ six members shall serve staggered two-year terms. No member completing a full term shall be eligible for immediate reappointment.

D. Publication Committee. The Publication Committee shall be constituted as described in ARTICLE 10 in the By-Laws.

E. Constitution Committee. Not more than eight years after the adoption of the Constitution and By-Laws, a Constitution Committee shall be appointed by the President for the purpose of reviewing the Association's Constitution and By-Laws, and preparing a revision if necessary, to be submitted to the membership not more than ten years after the adoption of this Constitution, in accordance with provisions and exceptions of ARTICLE XXII of the Constitution. In the event that a major revision of the Constitution and By-Laws is considered by the Constitution Committee, requiring a great deal of time, the Board of Directors may extend the life of the current Constitution and By-Laws for one year at a time.

ARTICLE 10. PUBLICATIONS

10.1 Editorial Boards. Each periodical published by the Association shall have an Editorial Board, consisting of all its Editors and such other personnel as may be designated by each Editor. The terms of Editors shall be determined by the Publication Committee.

10.2 Publication Committee. The Publication Committee shall consist of the Board-appointed Editors of each periodical published or co-owned by the Association, the ~~Secretary~~ Executive Director (ex-officio), and non-editor members of the Association at least equal in number to the number of editors. The non-editor members shall be appointed

by the President with the consent of the Board of Directors and one of these members shall be designated as Chair. The non-editor members shall be appointed for three-year terms, with one-third of these members retiring at the end of each year. Proposals for the publication by the Association of periodicals, directories, special reports, books or other professional material, or for publication in cooperation with other societies or organizations of such materials, shall be referred to the committee for its recommendations, prior to action by the Board of Directors. The Committee shall generally supervise the publication policy of the Association and make recommendations to the Board of Directors with respect to the editorial policy of the Association's various publications.

10.3 At suitable intervals, a directory of all classes of members, the Constitution and the By-Laws of the Association shall be published and sent to each member.

ARTICLE 11. RESOLUTIONS

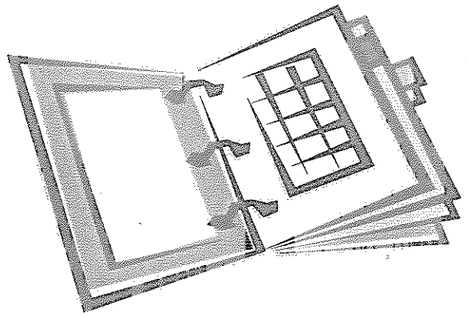
11.1 In no case shall the name of the Association be used in connection with any partisan or political issue, except insofar as the resolution of the Association refers solely to a matter involving the interest and objectives of the Association. Such a resolution shall require a favorable vote by at least two-thirds of the entire Board of Directors before it may be released for publication or transmission outside the Association. When there is doubt as to whether an issue is to be considered partisan or political, a majority vote of the entire Board of Directors shall be required to declare it not partisan nor political.

11.2 Resolutions and recommendations of Districts, Sections or Committees of the Association shall be so phrased as not to commit the Association or its membership.

ARTICLE 12. AMENDMENTS TO BY-LAWS

12.1 Proposal. Amendments to the By-Laws may be proposed by the Board of Directors or by a petition signed by at least 25 individual members. An amendment originated by petition shall be referred to the Board of Directors for its recommendations as to ratification.

12.2 Ratification. Following action by the Board of Directors, the ~~Secretary~~ Executive Director shall publish a copy of the proposed amendment and the results of the Board of Directors' vote on recommending ratification in the next issue of a newsletter, inviting comment. At least four weeks shall elapse between publication and the vote on an amendment. If, during this period at least 25 individual members of the Association so petition, the amendment shall be submitted to the individual members for a mail vote. A minimum of a two-thirds affirmative vote of the individual members voting shall be required for ratification. If no such demand for a membership ballot is received, the amendment may be ratified or rejected by the Board of Directors. Ratification shall occur whenever two-thirds of the members of the Board of Directors have submitted an affirmative vote either in person or received by mail within a period determined by the Board of Directors.



Organising/Programme Committee:

W.K. Li (Chairman)	F.W.H. Ho	L.K. Chan
X.L. Meng	W.S. Chan	K.W. Ng
J. Fan	H. Tong	K.T. Fang
L. Zhu	M.G. Gu	

Conference Secretariat:

A.C.S. Wong P.L.H. Yu

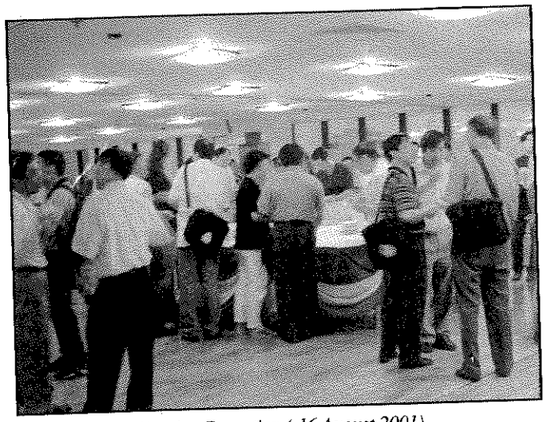
Report on the 5th ICSA International Conference

The 5th ICSA International Conference was held on 17-18 August, 2001 at the University of Hong Kong. As members can recall the first ICSA conference was also held in Hong Kong at the Chinese University on 15-17 December, 2000. Hong Kong is clearly an ideal and convenient meeting point for statisticians who were originally from the Southeast Asian region. The conference was declared opened in the morning of 17th August by Professor S.P. Chow, Pro-Vice-Chancellor of the University of Hong Kong, Professor Agnes Hsiung, President of ICSA and Professor Howell Tong. Professors Hsiung and Tong were the Conference Co-Chairs. There were altogether 280 participants from some 14 countries and regions, not including some 30 student helpers and un-registered graduate students. These include 65 from Hong Kong, 62 from the mainland, 32 from Taiwan, 76 from the U.S., 16 from Singapore, 13 from Canada with the rest from U.K., Australia, New Zealand, Japan, Macau, Germany, Denmark and Finland. For the first time, the conference was also co-sponsored by the prestigious Institute of Mathematical Statistics.

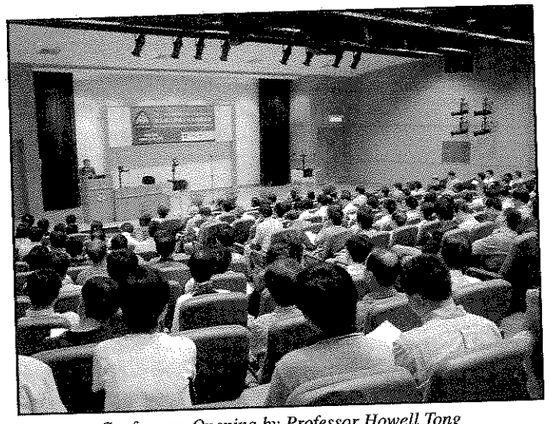
The academic programme had about 75 sessions and altogether some 200 papers had been presented. For a two and a half-day conference this meant that apart from the keynote addresses there were about 7 to 8 parallel sessions at any one time during the official hours of the scientific programme. The academic programme was also lucky to have two eminent statisticians as its keynote speakers namely, Professor Tze-Leung Lai and Professor Peter Hall. The abstract of their keynote speeches were attached for the interest of the members. In addition to the keynote speeches, we had also many interesting papers delivered by Professors Wing Wong, Kai-Tai Fang, Jia-An Yan, Jianqing Fan and David Siegmund, to name but a few. There was also a session on official statistics with speakers from the Mainland, Taiwan, Hong Kong and Macau.

On the lighter side, the conference was started with a reception on the evening of 16th August at the staff club of the University of Hong Kong. The conference banquet was held at the Miramar Hotel in the evening of the 18th August. There were some 23 tables and each participant was presented with a beautiful crystal dice as a table gift. The faces of the dice show the ICSA logo and the landmark of Central, Hong Kong.

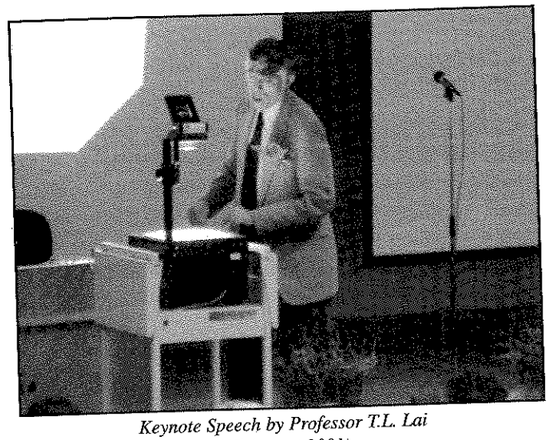
The programme ended officially by 12:30 on Sunday, 19th August. However, the Scientific Programme and photos of the conference can still be downloaded from the website, http://www.hku.hk/statistics/ICSA2001/ .



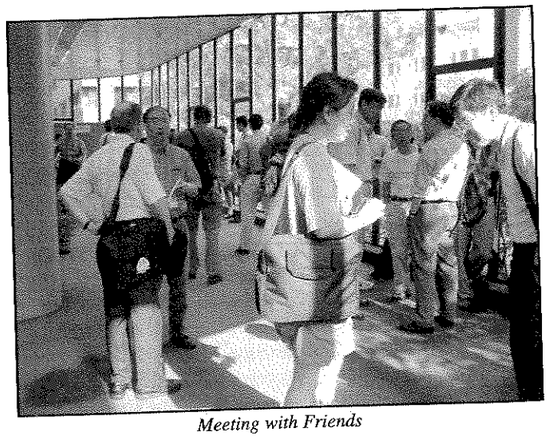
Evening Reception (16 August 2001)



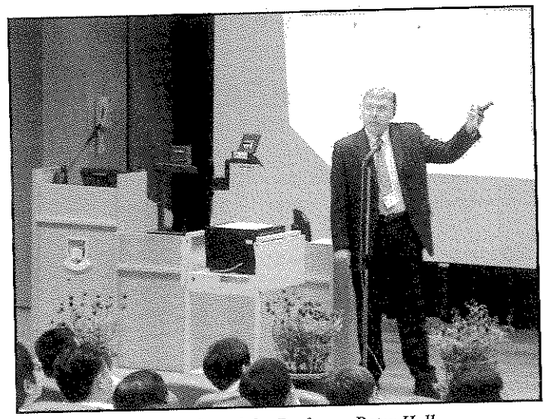
Conference Opening by Professor Howell Tong (Morning, 17 August 2001)



Keynote Speech by Professor T.L. Lai (17 August 2001)



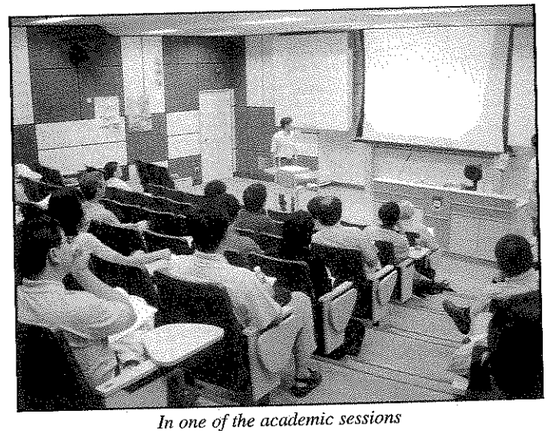
Meeting with Friends



Keynote Speech by Professor Peter Hall (Morning, 18 August 2001)

An Enjoyable Evening—the Conference Banquet (18 August 2001)

Conference Banquet (18 August 2001)



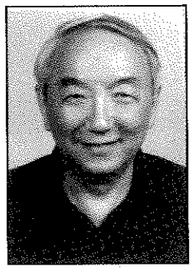
In one of the academic sessions

Interview with Professor George C. Tiao

By Ruey S. Tsay

Date: January 3, 2002

[Interview questions are provided by Sue-Jane Wang and William Wei]



Editor's note: The Editorial Board would like to thank Drs. Timothy Chen, William Wei, Sue-Jane Wang, Kung-Yee Liang, Yi Tsong, Grace Young, etc. for their enthusiastic support and effort during the preparation of the special interview

BACKGROUND

George C. Tiao is W. Allen Wallis Professor of Econometrics and Statistics, University of Chicago. Born in London in 1933, he returned to China with his parents in early 1934. He graduated from the elementary school in Chungking the year Japan surrendered. During the Chinese civil war, he moved with his parents and the Nationalist Government from Chungking to Nanking to Shanghai to Chungking to Hong Kong and finally to Taipei in 1950 and, as a result, attended a number of middle and high schools. He earned his B.S. in Economics from National Taiwan University in 1955 and came to the States in 1956. He earned his MBA from New York University in 1958 and his Ph.D. in Economics from University of Wisconsin-Madison in 1962.

George was assistant, associate, full, and Bascom Professor of Statistics and Business from 1962 to 1982 at the University of Wisconsin-Madison, and served as Chairman of the Statistics Department from 1973 to 1975. He joined the Graduate School of Business, University of Chicago in 1982. A leader in Bayesian inference, environmental study of stratospheric ozone and air pollution, and time series analysis, he has authored, co-authored and co-edited 8 books and more than 120 articles in leading econometric, environmental and statistical journals. He has supervised more than 25 Ph.D. and served in numerous academic committees in the United States, China, and Taiwan. He has received many honors, including the 2001 Wilks Memorial Medal and the Shiskin Award of the American Statistical Association, Fellow of ASA and IMS, and elected member of Academia Sinica, Taiwan in 1976.

He is the Founding President of ICSA and the Founding Chair-Editor of Statistica Sinica.

Q1: Many of our newer members are curious about the history of our organization. As one of the pioneers of the International Chinese Statistical Association, please tell us the early life of ICSA, for instance, from its creation to the toddler stage of early 90s.

George: It all started in Wisconsin. I first attended the ASA annual meeting in 1961 and to my surprise there were only three Chinese statisticians; two in economics and one in biostatistics. Things started to change in the late 60s and early 70s. My wife and I bought a house with a basement in Madison in 1967 and started to invite Chinese statistics students and their families to the Thanksgiving dinner at the basement. This "dinner festivity" grew to more than 80 participants and lasted for more than 20 years. Many students came to prepare the food the day before Thanksgiving and help cook dinner the next day. We have very nice memory about these dinners. As a matter of fact, many former students know my wife much better than me because of the dinner and opportunities to get together. It was at the 1968 dinner that I realized we need an association to promote more communication and collaboration among Chinese statisticians. We started with an informal association called the Chinese Statistical Society in U. S. With the help of 8 to 10 enthusiastic volunteering students, a hand-written bulletin that contains the directory of Chinese statisticians was published in the following year. The principal student leaders for the first two years were Austin Lee and Der-An Hsu. After the third year I thought that it is better for the Society to broaden its base by locating it around the States. I asked Professor Y. S. Chow of the Columbia University for help. He recruited Min-Te Chao, later the Founding Director of the Institute of Statistical Science, Academia Sinica, to be responsible for the administrative activities of the Society. At that time, Chao was at the Bell Lab. Without the help of graduate students, the publication of the annual bulletin quickly became an impossible burden for any single individual. Three years later, the Society moved back to Madison and was renamed Chinese Statistical Association in America.

Besides the bulletin, Chinese statisticians got together to have dinner each year at the annual ASA meeting as a means to get acquaintance with one another. As I recall, this started in the early 1970's at the St. Louis meeting with Hubert Chen as the first organizer. It began with 10 to 20 people, but grew quickly to more than 100 by the beginning of the 1980's. It also became a regular event with an informal meeting in the late afternoon of Wednesday followed by the dinner. With the expansion, we started to wonder whether such an arrangement is effective and sufficient for promoting the communication among Chinese statisticians. The final push to have a formal association has to do with Statistica Sinica.

Three key developments occurred in 1986. First, I returned to Taiwan for the Academia Sinica members meeting and had a chance to meet with Director Chao and other statisticians. Chao suggested that the Institute is sufficiently mature and has budget to launch a new statistical journal. We felt that to publish a new journal it is best to involve all Chinese statisticians inside and outside of Taiwan. Second, the 1986 ASA meeting was held in Chicago and Jia-Yeong Tsay suggested at the afternoon meeting that the time is ripe for the association to be formalized. He, Grace Yang and Gordon Lan formed a committee to draft the constitution of ICSA in a similar spirit to the ASA Charter. Third, at the dinner James Fu told me of his plan to launch a new statistical journal. I suggested to James and Min-Te that it is best to combine their efforts to establish a world wide first-rate journal. They came back to me about the Christmas time that year saying that they decided to cooperate and wanted me to formally launch the journal. To make the long story short, I consulted with several senior Chinese scholars including the late Professor Shein Ming Wu from Madison. They were all very supportive and gave me valuable suggestions including

having a strong local support at Chicago. As you know, Wing Wong was in Chicago then and he gave me his whole-hearted enthusiastic support. Ruey Tsay and Xiao-Li Meng also came to Chicago shortly after. Furthermore, I also obtained enthusiastic support from Smiley Cheng, T. L. Lai, L. J. Wei, and Jeff Wu. Thus, I decided to accept the challenge. An editorial board was formed in April 1987 and I served as the Chair-Editor. Because the Institute cannot sign an agreement with an informal association, the ICSA was formally established in 1987 to jointly sponsor the journal, Statistica Sinica.

The mission of ICSA clearly states that the association should not only provide services and communication among Chinese statisticians in North America, but also help promote statistical theory, application and education among Chinese communities in mainland China, Hong Kong, Singapore and Taiwan. To this end, one of the first goals of the Association when established was to hold an international conference in Hong Kong in 1990. The conference was organized by S. Y. Lee at the Chinese University of Hong Kong and was a great success, and subsequent international conferences held in Taipei, Beijing and Kunming were equally successful. Last year, the 5th international conference was again held in Hong Kong. Another success story of ICSA is the annual Applied Statistical Symposium. Many members of ICSA are biostatisticians working in the pharmaceutical industry and in the federal agencies such as FDA and NIH. Jia-Yeong Tsay and Gordon Lan initiated and organized a successful half a day meeting called ICSA Biopharmaceutical Statistics Symposium in Washington, DC in 1990, and it was expanded to include other areas beside biostatistics and changed into the current name two years later. This symposium has become an annual event and grown into a three-day affair attracting more than 200 participants from around the world.

Q2: You were the recipient of the prestigious Shiskin and Wilks Memorial Awards in 2001. Could you tell us what were the specific accomplishments cited in these two awards?

George: I was deeply honored to receive these two awards. For the Shiskin Award, they cited my contributions to the analysis of business and economic data in general and the model-based approach to seasonal adjustment in particular. For the Wilks Medal, it is the accumulation of my modest contributions to statistics. As stated in the criteria for the medal, Professor S. S. Wilks had made outstanding contributions in many areas of statistics including statistical theory, statistical application, and services to the profession. Thus, the Committee considered the contributions of recipients from all aspects. In my case, they cited my research contributions in Bayesian inference, environmental study on ozone depletion and air pollution indexes, and time-series analysis, my services to the profession in various committees and education, and my efforts in promoting statistical education in China and Taiwan.

Q3: You have been nominated to be a candidate for the ASA President in 2002 election. Congratulations! From the ASA perspective, where do you think the relationship between ASA and ICSA stands now, and where and how do you think the relationship between the two societies may be strengthened?

George: I am honored to be nominated to stand for the President of the ASA, which by all accounts is a very generous association to minority groups. For example, it has routinely provided a room for ISCA members to hold their annual members meeting during the annual joint statistical meetings for many years. The two associations have many common members and share many common goals such as promoting statistical education and research across different disciplines and different countries. Regarding the relationship, I would say that the two societies have already had a very good start, due to the efforts of a number of colleagues notably S. C. Chow, Tim Chen, C. P. Han, Huey Ju and several recent ASA Presidents such as Jon Kettenring. ICSA has in recent years played a more active role in the joint statistical meetings such as sponsoring an invited session to feature selected articles published in Statistica Sinica. In addition, several presidents of ASA have been invited to deliver keynote speeches at the ICSA Applied Symposium. For the future, there are many ways the two societies can further cooperate. I think the relationship between ASA and the Canadian Statistical Society is a good example to follow. ICSA can co-sponsor awards, the annual joint meetings, and other more focused conferences. For instance, I believe that the ICSA Applied Statistical Symposium can serve as an example for a winter or early spring meeting of ASA. The emphasis on substantive applications and short courses to disseminate state-of-the-art knowledge in statistics will distinguish such a winter meeting from the annual summer meeting. Furthermore, we can add a placement service for new graduates in the winter meeting as the time corresponds to the academic recruiting season.

Q4: As a candidate for ASA President, can you tell us about your vision and mission for this endeavor?

George: Thank you for having the opportunity to share my vision and mission with members of our association. This is an exciting time for statisticians and an important time for the statistics profession. We are witnesses to and participants in a worldwide explosion in the collection, manipulation and analysis of data addressing issues of societal importance. Yet statisticians do not command the attention and respect of the society as we should. Our profession is at a critical juncture and associations such as ASA and ICSA must make good use of the opportunity to expand the influence of our profession. If elected, I would focus on two main directions. First, I would take concrete steps to expand the role and the visibility of statistics in science, industry, business, and public policy. Second, I would move aggressively to unify the profession across diverse statistical disciplines and identify substantive ways to shape the ASA into an organization that (i) furthers the growth of the profession, (ii) promotes the development of statistical theory and methods in individual application areas, and (iii) fosters inter-disciplinary exchanges that strengthen the profession as a whole. More specifically, as shown in my statement to ASA, I will give priority to the following areas.

New challenges in emerging areas.

There are great opportunities in information technology, life sciences and biopharmaceutical research. Building on the ASA's long tradition of cooperation with sister statistical associations, we should forge strong links with other societies (e.g., ACM, IEEE, DIA, etc.) by co-sponsoring meetings, revising curriculum, forming lobby groups, and participating in the drawing of funding priorities and government policies. Efforts of this kind will stimulate collaborative projects leading to advancement of statistical theory, methods and application.

Proactive role in federal agencies.

ASA should play a key role in assisting government agencies such as FDA and NIH to establish guidelines/guidance for drug research and development. In addition, for agencies such as NASA and EPA, which generate enormous amount of scientific data but do not have central statistical offices, ASA should take the lead in persuading the government to create individual advisory committees for overseeing statistical functions and dissemination of quantitative information. To help achieve this goal, I will seek to form an ASA standing committee consisting of prominent colleagues with established reputation in specific application areas. Members of the committee can then be called upon to speak for the profession on issues of critical importance and to advise government agencies on statistical functions.

"Professional" degrees in Statistics.

Practicing statisticians will command more attention and compensation if their degrees are regarded more like

professional degrees such as the MBA, MPA or MPH. The ASA should take the lead in promoting the creation of these types of professional statistics degree programs. They can be two-year programs for a MS in Statistics/Biotechnology, Statistics/Environment, Statistics/Financial Economics, Statistics/Law and the like. Excellent examples of MS programs with an applied concentration have already been established at Carnegie Mellon, Columbia, Iowa State, NC State and a number of other major statistics departments. Under ASA sponsorship and in collaboration with professional experts from individual substantive disciplines, standardized curriculum and test can be established to help maintain their quality. Thus, I believe that these kinds of degree programs, if successfully implemented, will greatly enhance the marketability of well-trained statisticians and lead to sustain long-term growth of the profession.

Enhanced functions of chapters.
ASA chapters are its grass root organizations. They are the best communication channels for academia, industries and government agencies in the same region. Thus chapters provide great potentials for collaborative research and development activities among the three main branches of our profession. There should be a membership drive to increase dues-paying members of the chapters. The ASA office should also provide greater resources to promote collaborative projects in the chapters and facilitate networking among chapters.

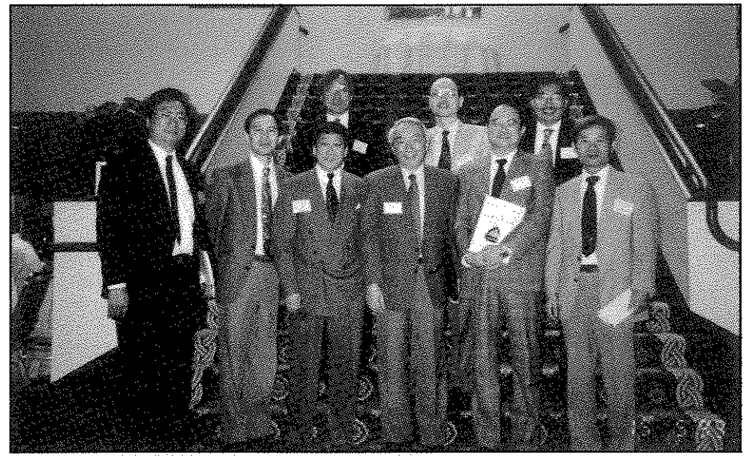
International collaboration.
In the age of globalization, ASA should play a more active role in the international arena. It can co-sponsor theme conferences with societies in other continents, and increase membership in other countries. Through collaboration with international and individual country statistical organizations, it can foster an international network of statisticians with interests/expertise in any given area of emerging importance.

Membership service.
ASA should continue to increase opportunities for members to obtain additional statistical training and learn new applications. It should also explore the possibility of a more focused winter or spring meeting that emphasizes applications and short courses, and provides a placement service that would more conveniently coincide with the academic hiring season. This second meeting could possibly be held jointly with another society or led by sections / chapters of ASA.

Having stated my own vision as a candidate of ASA President, I would like to return to your previous question regarding the cooperation between ASA and ICSA. It is my sincere belief that members of both associations can and should work closely together to accomplish these goals, which, in my opinion, would in turn strengthen our profession as a whole in the long run.

Q5: You have successfully spread the influence of statistics worldwide. Your efforts in Asia such as China, Hong Kong, Singapore and Taiwan are particularly noteworthy. As many of our members are seeking guidance regarding their career choice and development, what would you like to tell us regarding your experience and any words of advice?

George: This is an important question, and I can only provide some thoughts for our members to consider. First, I am a strong believer that statistical theory and application must go hand in hand. This was true in the past and will be even more so in the future. Advances in science and technology have opened new frontier for statistical thinking and applications. We must prepare ourselves for the information age. Having professional degrees in statistics mentioned in my mission statement is just one of many ways to prepare our profession for the future. The degrees emphasize on statistical training as well as learning knowledge of the specific area of application. We must understand the substantive characteristics of the problem under study for statistical theory to be relevant and statistical application to be successful. Second, a specific yet related example of a most promising and important application area is biostatistics and bioinformatics. As the chair of the advisory committee for the Division of Biostatistics and Bioinformatics, National Health Research Institutes, it is clear to me that Taiwan has had a good start, even though it still has a long way to go. China, on the other hand, is a vast, virgin land! In the midst of its phenomenal economic growth, how to address the urging issues in public health and medical advancement poses a serious challenge the resolution to which our profession and in particular, our members, can and should get involved. Members interested in the area must learn statistical theory and methods and, in addition, take relevant courses in biomedical science, computing and even humanity. Finally, I would like to thank you, Bill, and Sue-Jane for your effort and the interview.



Taken with ICSA members at an Applied Statistics Symposium

Taken with a group of colleagues at the 4th ICSA International Conference, Kunming during a tour of the West Mountain of China, 1998

SPECIAL TOPIC

PHARMACOGENOMICS

In this special topic section, we will introduce use of statistics in pharmacogenomics ranging from what is pharmacogenomics "Challenges and Opportunities for Pharmacogenomics" by Frank Shen, to targeting single gene to multi-locus genes "Statistics Applied to Classical Genetics – Finding Disease Responsible Gene(s)" by Sue-Jane Wang, followed by "On the Potential of Tree-based Analyses in Genomics" by Heping Zhang, and "Is Statistical Significance still Significant in the Post Genome Era?" by Lue-Ping Zhao.

Editor: Sue-Jane Wang, Ph.D.

Challenges and Opportunities for Pharmacogenomics

C. Frank Shen, Ph.D.
Biostatistics and Data Management for Clinical Discovery
Pharmaceutical Research Institute
Bristol-Myers Squibb Co.

The Time is Christmas Eve of 2020. Mr. Gene Snip and his family are getting ready to go to the hospital to visit his father who was diagnosed with late stage prostate cancer six months ago. Before they leave the house, Gene quickly gets on the internet and checks his e-mail. There were three e-mails sitting in the inbox. The first one is from the Genetic Trust Bank, where he deposits his blood sample, to inform him that there were two research institutes and one pharmaceutical company interested in obtaining his blood sample for their genetic research and to ask for his consent. The second one from the Lotus Genetics Ltd. returns his genetic testing results that he ordered on-line from the Genetic Trust Bank website two weeks ago. In that e-mail, they tell him that he does share the polymorphism that caused his father's cancer. The third e-mail is from the RiteGenetics Pharmacy website to inform him about personalized treatments available to prevent his cancer.

This is the classic picture of future preventive and personalized medicine. Pharmacogenomics is a new field on the horizon that can paint the picture. It has attracted talents from multiple disciplines: biologists, geneticists, medical doctors, biostatisticians, epidemiologists, computer scientists, who all roll up their sleeves to solve this unprecedented puzzle together. It is well known that some drugs work better in some patients than in others, and some drugs may even be highly toxic to certain patients. Pharmacogenomics is about spotting correlation between such responses to drugs and the genetic profiles of patients. It generates data that are relevant to a drug's clinical information. The genetic basis for drug response is quite different from the more traditional disease genetics and can be segregated into three categories: (1) pharmacokinetic dimension such as bioavailability related to absorption, disposition, metabolism, and excretion (ADME); (2) pharmacodynamic dimension which is the interaction of the drug with its target, whether it be a receptor or an enzyme; (3) the overall issue of side effects. For pharmaceutical companies, Pharmacogenomics is the use of genetics to optimize drug discovery and development.

It is perhaps the most exciting journey in human history that current medical practice based on drug selection can be substituted by patient selection, but clearly there is a long way to go from 2001 to 2020 or even beyond. Challenge and opportunity ahead of Pharmacogenomics are like two sides of the same coin. The first challenge is in the functional genomics that ties the gene name with its disease target (i.e., function). Very few genetic markers have been validated to date as indicators of drug response. Finding these markers is still largely a hit and miss process and we are still scratching the surface in assessing the clinical impact of genetic markers. More advances in bioinformatics are required to analyze enormous amounts of data due largely to the advent of DNA chip technology and the automation of DNA cloning and sequencing that led to the completion of the human genome project.

Although we are getting better everyday in building maps for functional genomics, our challenge does not stop there. Human are far more complex than a series of changes in a linear genome that Pharmacogenomics may pick up. There are strong environmental influences as well as lifestyle and other factors that influence how much drug is available at any given time. We are different ages, different health, different diets, and of course, we take different drugs. Those are enough differences that have nothing to do with individual's genetics. Sounds hopeless? Not quite. As always in genetics you study twins when you want to see if something has a strong genetic background versus environment. Some unjustified twin studies from United Kingdom have shown that the contribution of genetics to drug response was about 80-85%. This means that only around 20% of the variation in drug response may be accounted for by environment.

Even as the dust settling and we are confident that we can develop drugs

that will differentiate response, taking it into clinical trial could be another daunting task. Imagine a genetic map that has as many as 100,000 genetic markers on it. Going into a single clinical trial with 2,000 people would require assay 200 million genotypes. This dwarfs most genetic analysis that has been done to date. To do just the 100,000 genotypes and PCR reactions individually would cost about $20,000 and 10 days assay time per patient using today's technology. Until we can push the cost down to $200 and time down to a couple hours per patient, the obstacles will prohibit obtaining genotypic information from a typical clinical trial routinely. We need new, improved technology.

Pharmacoeconomic considerations are as important as scientific and technical considerations. It is advantageous to use pharmacogenetic profiling to understand the patterns of responsiveness and non-responsiveness based on inter-individual genetic variation. For example, drug-drug interaction is the fourth largest cause of death in the United States and much of that relates to genetic polymorphism in the metabolic pathways. But which genetic testing is the most economically feasible? I would think for drugs that are on the market, highly priced drugs with high non-responsive fractions (e.g., -interferon) are the most obvious candidates while high-volume, low-cost agents (e.g., beta-blockers) seem less viable candidates for Pharmacogenomics. Similarly, some disease areas with higher genetic relevance such as oncology or anti-infective are much more applicable than others such as cardiovascular or metabolic disorders, which is clearly more likely to be environmentally influenced. It would highly depend on each pharmaceutical company's portfolio and require further Pharmacoeconomic modeling to decide the marketing impact of Pharmacogenomics.

Last but not the least, Pharmacogenomics challenges regulatory agencies for a new paradigm for drug approval. For example, future pharmacogenomic-based drugs are likely to be marketed together with genetic testing for screening patients and would require an integrated review. On the other hand, data is the core of drug approval. Despite a great deal of discussion as to the huge potential of genetic information in development, very little hard data is available. Until companies work closely with regulatory bodies to work out where genetic data is going to be useful, regulatory agencies will continue to view pharmacogenomics as a tool for hypothesis generation and genetic variation as just one of many factors that contribute to drug response.

Like many new technologies, it takes time to sort out problems in the dark. But once it comes out of the dark, it will travel at a quantum speed. Recent computer development is a good example. It took about 10 years to go from a 286 to a 486 computer processor, but only a fraction of that time from 486 to Pentium IV. Those tough challenges in Pharmacogenomics present the best opportunity to stimulate the collaboration of different fields working together and produce the best hope to make drugs better targeted, safer, and more efficacious.

Statistical Methods Used in Classical Genetics Potential Identification of Disease Responsible Gene(s)

Sue-Jane Wang, Ph.D.
Division of Biometrics II
Office of Biostatistics
Center for Drug Evaluation and Research
U.S. Food and Drug Administration

Prior to Joining FDA
Quantitative Mathematical Genetics and Epidemiology
Division of Medical Genetics
Department of Pediatric
Cedars-Sinai Medical Center
Los Angeles, CA, USA

The definitions of interest in classical genetics predate the modern molecular era. To learn more about statistics tools utilized in classical genetics, first, I would like to introduce some indispensable terminology to facilitate understanding.

A gene is a specific coding sequence of DNA (Elandt-Johnson 1971), the unit of transmission, recombination, and function (Vogel and Motulsky 1986). Heritable characters are determined by genes, where different genes are responsible for the expression of different characteristics. Genes occur at definite sites, or loci (a collection of genes) along a chromosome. Each locus can be occupied by one of several variant genes called alleles. For instance, the ABO locus residing on the long arm of chromosome 9 at band q4 has three alleles: A, B, and O. Pairs of alleles constitute a genotype, which may not be observable. What is observable is a person's phenotype. In the ABO system, there are six unordered (not distinguishing the parental source) genotypes: AA, AB, AO, BB, BO, OO, and there are four phenotypes: A, B, AB, and O. Alleles A and B are co-dominant alleles and allele O is a recessive allele.

In this Special Topic column, I will introduce statistics methodologies often used in classical genetics. Specifically, I will focus on the following three topics: Hardy-Weinberg law, segregation analysis and linkage analysis.

The Hardy-Weinberg Law (or Hardy-Weinberg Equilibrium, HWE) of population genetics allows estimation of gene frequencies from allele frequencies. The mathematical model to establish HWE has seven explicit assumptions: infinite population sizes, discrete generations, random mating, no selection, no migration, no mutation, and equal initial genotype frequencies in the two sexes. HWE is equivalent to the random union of two gametes, one gamete being an egg and the other being a sperm. Using the ABO example, let p_A and p_B be the allele frequencies for alleles A and B, respectively. Then a random individual will have phenotype AB with frequency $2*p_A* p_B$. The factor 2 reflects the

two equally likely ordered genotypes A/B and B/A. Notice that the ordered genotypes preserve the inheritance phase. That is, the inheritance phase defines a particular allele's parental source, paternal or maternal. The HWE is somewhat subtle for X-linked loci because twice as much weight is attached to the initial female frequency since females have two X chromosomes while males have only one. Therefore, the frequency in males is always the frequency from the preceding generation and the frequency in females is the average frequency for the two sexes from the preceding generation.

Rigorous analysis of human pedigree data is a vital concern in genetic epidemiology, human gene mapping, and genetic counseling. Because of the complexity of genetic models, statistical methods applied are generally the maximum likelihood estimators. Special techniques such as simple counting arguments and the EM algorithm appear to be very appealing for numerically computing maximum likelihood estimates.

The objective of classical segregation analysis is to test Mendelian segregation ratio in nuclear family data. A nuclear family consists of two generations. Segregation ratio is the proportion of outcome for a specific cross. This proportion is unambiguous if the transmission of gametes follow the Mendelian law of autosomal dominant, autosomal recessive, or autosomal co-dominant forms. In these cases, a homozygote individual with the genotype A1A1 produces only A1 gametes to their offspring. A heterozygote individual with the genotype A1A2 yields gametes A1 and A2. Thus, the cross A1/A2 by A1/A2 produces offspring in the ratio of ¼ homozygote for A1/A1 and A2/A2, respectively, and ½ heterozygote A1/A2. These proportions are segregation ratios. Usually the hypothesis of interest is that some rare disease shows an autosomal recessive or an autosomal dominant pattern of inheritance. When the disease is rare, it is inefficient to collect families at random. The approach taken in the genetics study is often to ascertain families with at least one affected sibling. Hence, to test the Mendelian segregation ratio of $p=1/2$ for an autosomal dominant disease or $p=1/4$ for an autosomal recessive disease, one must correct for the ascertainment process. One simple ascertainment model assumes that the number of ascertained siblings follows a binomial distribution with success probability π and number of trials equal to the number of affected siblings. The number of affecteds (siblings) follows a binomial distribution with success probability p and number of trials equal to the number of siblings. This is because families are ascertained only through their affected siblings, and the siblings have a common probability of success π. The EM algorithm can be employed to estimate π and p jointly.

When the mode of inheritance cannot be clearly stated to be autosomal dominant, recessive, or co-dominant, the mode of inheritance would rely on the penetrance parameter. To analyze human pedigree data is tedious without computer's help. Pedigrees lack symmetry and all simple close-form solutions in mathematics depend on symmetry. In statistical genetics, one major advance in the literature is the pioneering algorithm of Elston and Stewart. Their contribution is to recognize that closed-form solutions are less relevant than good algorithms. The likelihood representation of a Mendelian model consists of three parts. One is prior probability of a genotype the founder carries. A founder is the top generation individual in the pedigree that does not have ancestor to trace further. Penetrance probability is another component of the likelihood function. Penetrance functions specify the likelihood of an observed phenotype given an unobserved genotype, i.e., pr(y|g). Given individuals' genotypes (g), their phenotypes (y) are independent. It is known that not all diseases are genetic. A model to incorporate the environmental factors would be desirable, that is, a polygenic model, which takes into account both the genetic and the common environmental influences on the expression of phenotypes. The third component is the transmission probability. Transmission of the trait or traits observed is the probability that a mother with genotype G_m and a father with genotype G_f produce a child with genotype G_c. It is calculated by multiplying two independent transmission probabilities of gametes from mother conditioning on mother's genotype $pr(G_c \mid G_m)$ and that conditional on father's genotype $pr(G_c \mid G_f)$. Classical likelihood function takes summation over all possible genotypes of all individuals in the pedigree on the product of "joint probability of transmission probability on the vector of all phenotypes conditional on all genotypes" and "prior probability on the vector of all individuals' genotypes." This joint probability is computationally prohibitive. Elston and Stewart's algorithm takes apart the transmission vector probability into the product of individual penetrances and the prior vector probability into product of founder's prior and non-founder's transmission. Computational efficiency dramatically improves. Of course, evaluation of the pedigree likelihood remains in need of further theoretical improvement.

Another area of application is linkage analysis. That is, identifying whether there is disease responsible gene(s) causing the risk of an offspring's disease. One of the first questions to be asked regarding the transmission of a disease is: Is it genetic? Sometimes, a concentration of disease cases among family members is not specific for genetic factors, e.g., infection. A main clue for an etiologic involvement of genes stems from the observation that the disease "runs in families." Linkage calculations introduce the linkage parameter θ in the likelihood construction. The parameter θ measures the probability that alleles locate on the chromosome travel together with the disease gene, which runs in families. Statistical techniques used in current linkage analysis are mostly based on maximum likelihood estimation and likelihood ratio testing. Instead of investigating a specific marker locus, multipoint linkage analysis explores linkage of a genetically inherited disease with several loci. Until recent literature, Bayesian principles are incorporated in multipoint linkage analysis. Others such as association between phenotypes and alleles are often of interest, too.

Classical linkage analysis can deal with a few loci simultaneously. With the advance of single nucletide polymorphisms and microarray technologies, small scale of the number of loci used to evaluate linkage analysis is gradually replaced by the parallel techniques. Microarray technology allows tens of thousands of genes be evaluated simultaneously. Most of the microarray work is to identify gene profile for subgroups such as kidney failure rats, lung cancer patients, etc.

Interested readers are referred to a few references for further insight on statistical applications in this fascinating field.

Some References

Elston R.C. and J. Stewart. 1971. A general model for the analysis of pedigree data. Human Hereditary. 21:523-42.

Methodology in medical genetics. An introduction to statistical methods by Alan E.H. Emery. Churcuil Livingstone, Longman Group Limited, 1986.

Analysis of human genetic linkage by Jurg Ott. The Johns Hopkins University Press, Baltimore and London, 1995.

Bioinformatics basics application in biological science and medicine by Hooman H. Rashidi and Lukas K. Buehler. CRC Press, 2000.

Some references on microarray related topics would not repeat and can be found in the articles by Heping Zhang and Lue-Ping Zhao

**The views expressed in this article do not necessarily represent those of the U.S. Food and Drug Administration*

On the Potential of Tree-based Analyses in Genomics

Heping Zhang, Ph.D.
Yale University School of Medicine
Email: Heping.Zhang@Yale.edu

Numerous human diseases are believed to have a genetic component, and in most cases, the underlying genetic mechanisms are likely to be complex. While traditional association and linkage analyses continue to be valuable in identifying candidate genes for diseases, they face increasing challenges.

The Human Genome Project, in an effort to sequence the entire human genome of approximately 30,000 genes, is near its completion and has opened different venues not only for identifying candidate genes, but also for understanding genes on pathways relevant to disease development. The information generated from the Human Genome Project has led to the emergence of a relatively new, powerful technology – the so called gene chips or microarrays. This technology enables us to simultaneously monitor the expression profiles of thousands of genes with data from a single slide or image. As a result, an explosive amount of information has been generated. Remarkable progression has been made in understanding such data and the technology has shown great promise in advancing our scientific knowledge. Nonetheless, one of the bottleneck problems in processing and exploring these data is the analytic tools and statistical inference.

In statistics, we typically deal with a small number of variables, and the properties of the statistical procedures usually rely on a large sample. However, gene chips produce the opposite data scenario where we have a large number of variables (i.e., the number of genes on a chip) and a small number of samples (e.g., the number of tissues under a particular study). Thus, it is not surprising that most of the existing statistical analyses in this area are explorative in nature.

Although most of the microarray data have been analyzed by using various clustering algorithms, in this article, I will briefly introduce another type of statistical methodology (Breiman et al., 1984 and Zhang and Singer, 1999) in genomic applications, namely, the tree-based methods. Tree based methods are nonparametric statistical approaches that can be used for conducting regression and classification analyses. They are very flexible for dealing with a large number of (possibly correlated) discrete or continuous variables, making them particularly appealing for microarray data analyses. Both commercial and free software (e.g., my RTREE from http://peace.med.yale.edu) are available to conduct such analyses.

The thrust of the tree-based methods lies in the so-called recursive partitioning technique.

Suppose we have data from a number of observations, each of which contains a vector of measurements or covariates (gene expression profiles from a tissue) and a class label (normal versus tumor, reference versus target). The aim of recursive partitioning is to build a classification rule that predicts the class membership based on the feature information. This technique normally uses one covariate and splits a study sample into two subgroups at a time, in order to find two relatively homogeneous subgroups judged by the mixing of the class labels within each subgroup. The most desirable subgroup is the one with only one label, although we also need to keep in mind the size of the subgroup, because a pure but small group is not as informative as a large homogeneous group. At the end of the recursive partitioning process, we hope to be able to extract homogeneous strata of the original sample, within each of which the class label is homogenous enough to provide us with a quality and reliable class prediction for a future sample of the same feature while unknown class. Rather than explaining the technical details and terminology of tree construction, I now present a specific genomic application and focus on understanding and interpreting a resulting tree.

Several studies have used microarrays to analyze gene expression in colon, breast and other tumors and have demonstrated the potential power of expression profiling for classifying tumors (Hedenfalk et al, 2001, Alon et al, 1999). Gene expression profiles may offer more information than classic morphology and provide an alternative to morphology-based tumor classification. Using recursive partitioning, Zhang et al. (2001) analyzed a data set from the expression profiles of 2,000 genes using an Affymetrix oligonucleotide array in 22 normal and 40 colon cancer tissues published by Alon et al (1999).

For example, the following classification tree that divides the 62 tissues into four final groups labeled as Nodes 2, 5, 6, and 7. The first division of the 62 tissues was done through the expression level of the IL-8 gene. This gene was found correlated, in the medical literature, with the stage of colon cancer when the migration of human colonic epithelial cell lines and metastasis of bladder cancer occur. Then, the CNAX gene is used to split one of the resulting group, labeled as node 3. This molecular chaperone gene was found to have decreased expression in HT-29 human colon adenocarcinoma cells and involved in apoptosis in human prostate epithelial tumor cells. Finally, RAB3B, a member of the RAS oncogene family, is used to split node 4. It is associated with significant increase of the mRNA expression in a human leukemia cell line.

Two of the final groups (Nodes 2 and 7) contain 21 normal tissues and no cancer tissue. In contrast, the other two nodes (Node 5 and 6) contain 40 cancer tissues and 1 normal tissue. If we predict the tissue type at a node uniformly by the majority vote, i.e., a node containing more normal (cancer) tissues than cancer (normal) tissues is labeled as "normal" ("cancer"), we misclassify one normal tissue out of the total 62 tissues (error rate 1.6%). This is far more precise than the classification rules using other methods.

Without adjustments, our naïve assessment of tree predictive precision is likely to be overly optimistic. In addition, the correlation among many gene profiles makes it likely that there exist other competing tree structures that could have a similar predictive precision. The small number of tissue samples makes it even more urgent to evaluate the quality and variability of the tree. We refer interested readers to Zhang et al. (2001) for discussions on how to obtain a more realistic assessment. Overall, this example as well as the analyses of other existing data sets suggest the great promise of the tree-based methods in classification using microarray data. Not only does the tree possess a high predictive precision, but also there is supporting evidence that the identified genes are related to colon cancer -- the outcome of interest.

References

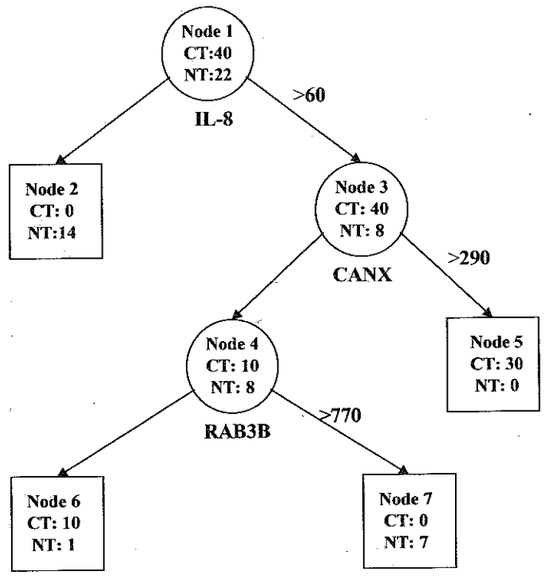
1. Alon, U., Barkai, N., Notterman, D. A., Gish, K., Ybarra, S., Mack, D. and Levine, A. J. (1999) Proc. Natl. Acad. Sci. USA 96, 6745-6750.

2. Breiman, L., Friedman, J., Stone, C., and Olshen, R. (1984) Classification and Regression Trees. Wadsworth, California.

3. Hedenfalk, I., Duggan, D., Chen, Y., Radmacher, M., Bittner, M., Simon, R., Meltzer, P., Gusterson, B., Esteller, M., Raffeld, M., Yakhini, Z., Ben-Dor, A., Dougherty, E., Kononen, J., Bubendorf, L., Fehrle, W., Pittaluga, S., Gruvberger, S., Loman, N., Johannsson, O., Olsson, H., Wilfond, B., Sauter, G., Kallioniemi, O.-P., Borg, A., & Trent (2001), J., N. Engl. J. Med. **344**, 539-48.

4. Zhang, H.P. and Singer, B. (1999) Recursive Partitioning in the Health Sciences, Springer, New York.

5. Zhang, H., Yu, C-Y, Singer, B., Xiong, M. (2001), Proc. Natl. Acad. Sci. USA 98, 6730-6735.



Is Statistical Significance still Significant in the Post Genome Era?

Lue Ping Zhao
Fred Hutchinson Cancer Research Center
Seattle, WA

Statistical significance has been a useful convention to assert investigative evidences in biomedical research. The concept of statistical significance is simple and intuitive. Formally, statistical significance is defined through a hypothesis testing framework, which is comprised of a null hypothesis, a test statistic measuring departures from the null hypothesis, and measurement of the chance that the test statistic suggests a departure from the null hypothesis, that is, the statistical significance. Since the inception of statistical research, a great deal of professional effort has been devoted to an appropriate evaluation and calculation of statistical significance in many scientific disciplines, including biology and medicine. Typically, statistical significance is measured by P-value, which takes on values between 0 and 1. A P-value of 0.01, for example, implies that under the null hypothesis, the chance that the test statistic takes a value equal to or greater than a critical value is approximately 1%. Consider a clinical trial as a practical example. Suppose that the drug being tested is ineffective, that is, the null hypothesis is true. The P-value measures the chance of incorrectly drawing a conclusion to the contrary, that is, the drug is effective. The erroneous conclusion is known as a false positive error or type I error, and is generally controlled at the 5% or 1% level. In spite of the technical and specific definition, statistical significance and P-value are frequently quoted, often incorrectly, in media, though sometimes incorrectly. It is one of the key parameters for regulatory agencies, such as the Food and Drug Administration (FDA), to make an objective assessment of drug effectiveness in making a decision to approve the drug for public consumption. Given the long-standing popularity of statistical significance, one may naturally accept the continued use in biomedical research, and would wonder why I write this assay about using statistical significance in the Post Genome era.

Before answering this question, I must define what I mean by the Post Genome era. In the early 1980's, several human geneticists realized the possibility of creating a comprehensive map of all known genetic markers (1). This prospect piqued the interest of several biomedical researchers, eventually leading to the development of the Human Genome Project (3). The initiation of the Project marked the beginning of the Genome era. Now 10 years later, the public consortium, later joined by a private effort, produced a rough human genome map (4, 7). The publication of the Map marked the ending of the Genome era, and the beginning of the Post Genome era. Through the Human Genome Project, biomedical researchers have learnt that the human genome has learned 3.2 billion nucleotides, which encode 30~40 thousand functional genes and over 100 thousand proteins. In addition, biotechnologists have developed numerous high throughput tools to assess variations in the human genome, gene expressions in cells, and protein expressions in tissues. For the first time in biomedical history, medical sciences have been transformed from an "information starved" to an "information rich" discipline. This is expected to revolutionize biomedical research in this Post Genome era, and success from genomic research will transform future health care practice in years to come.

As genomic technologies are increasingly applied to medicine in this era, the issue facing many statisticians, the topic of

this essay, is whether or not statistical significance is relevant. Indeed, the formal statistical approach is not commonly practiced in molecular biology, neither is the concept of the statistical significance. The primary reason is that molecular biologists are often able to validate their findings via repeating experiments in a timely fashion or via assessing findings with different technologies. Such validation experiments can be repeated cost-efficiently. This is in sharp contrast to the common practice of statistics in medical research, where validating a positive finding typically involves separate human studies and takes a significant amount of human and financial resources. Scientific evidences are typically measured via statistical significance. When molecular technologies are increasingly applied to medicine, one would naturally wonder if statistical significance will remain significant in this new era.

To assess the use of statistical significance in current literature, I performed an admittedly simple and surely incomplete search on PubMed, a free public database of journals (http://www.pubmedcentral.nih.gov), using "statistical significance" as a key phrase and then combining it with another key word/phrase. With "statistical significance" as the only key phrase, I found 13,745 entries on January 2, 2002 database. Then, I used several secondary key words/phrases more closely related to applied/medical research, including "human", "animal", "biomarkers", "population", "epidemiology", "medical", "clinical" and "pathology". Frequencies of entries from these searches are listed in the following Table 1. Also listed are numbers of entries by these key words/phrases themselves, useful as denominators. As expected, statistical significance is frequently used in applied/medical literature, ranging from the maximum frequency of 11,348 entries associated with "human" to the minimum frequency of 616 entries associating with "biomarker". In contrast, when combining "statistical significance" with a key word/phrase from basic science research, including "biology", "molecular biology", "cellular biology", and "biochemistry", the occurrence of "statistical significance" is much lower than those in medical literature, with the maximum frequency of entries 321 associating with biochemistry to the minimum frequency of entries 0 associating with molecular biology. Then, I searched for "statistical significance" with key words/phrases from genomic research, including "genomics", "functional genomics", "proteomics", "SNP", "GeneChip", "microarray" and "array". Not surprisingly, the occurrence of "statistical significance" was even lower. Admittedly, the low frequencies at this time are also associated with the recent development of relevant disciplines.

There are several reasons for the virtual absence of "statistical significance" in genomic literature. First, most molecular biologists have traditionally relied on experimental methods to validate any positive leads, and do not use statistical approaches in general. Because of their focused experiments, biologists can often rely on visualizations to reach conclusions, without performing any significance analysis. This tradition may have kept many genomic researchers from using statistical significance. Secondly, a small group of biologists, who actually have knowledge about statistics, may believe that the hypothesis testing framework is not particularly useful, since most genomic research is oriented towards discovery and is hypotheses-generating, rather than hypotheses-testing (2). Hence, statistical approaches are seen to be less relevant. Lastly, the concept of statistical significance is not well understood and many existing statistical methods are not directly applicable to genomic research. In a typical genomic study, researchers usually gather high throughput data on many genes from relatively few independent samples. In contrast, traditional studies often involve many independent samples with few observations on each sample. Thus, typical methods of evaluating statistical significance may not be directly applicable without careful scrutiny.

Despite the sensible reasons stated above, the statistical approach is still applicable to genomic data analyses and therefore, statistical significance should remain important as a key measure of

investigative evidence. First of all, the hypothesis-testing framework has been a foundation of biomedical research for many decades, and has been particularly useful for initiating and planning costly studies. Before launching into any biomedical study, investigators need to carefully choose hypotheses (or scientific questions), and assess the feasibility of completing the study with the desired power at a designated statistical significance level. Having an objective measurement such as statistical significance would help researchers to design an appropriate study. Secondly, relating to the first point, having a hypothesis framework also helps investigators to choose an appropriate study population in addressing scientific questions. The choice of a study population also imposes the constraints on generalization of results and relevance of statistical significance to other populations. Thirdly, even in "opportunistic studies" (based on available biological samples from historical studies), one needs to use statistical significance to measure investigative evidences. Additional caution is required, of course, when using historical data, such as the ways the data are gathered, limitations inherent to the sampling scheme and research questions to be addressed. Incorporating all the sampling constraints into statistical methodology, one should be able to carry out rigorous statistical analyses and to assess statistical significance and address questions of interest. I grant that the interpretation of statistical significance in hypotheses-generating studies may be not as rigid as in that in hypothesis testing studies. Nevertheless, naively mining data, out of context, could generate misleading conclusions. Further, if only visualization tools or cluster analyses are used for data mining, the subjective element inherent in these techniques would lead to unnecessary "beauty contests", without substance.

To ensure an appropriate adoption of statistical significance in genomic studies, we, as statisticians, need to take on a few of the tasks discussed below. We must learn how to communicate effectively with researchers in basic sciences. To achieve this objective, we must become familiar with modern biology and related concepts, including the human genome map, genes, transcripts, proteins, regulatory circuitry and microarray technologies. Meanwhile, we should also learn how to convey statistical concepts and ideas in context and in the simplest terms possible to our biological collaborators. Of course, one of the first concepts is statistical significance. Secondly, we must think critically about established statistical paradigms and methods of calculating P-values, as well as about alternative measurements of statistical significance to P-values. As discussed above, genomic studies have unusual features that are different from traditional statistical applications. Namely, a typical genomic study has few samples and many genes observed on each sample. This feature has immense implications in the calculation of P-values. Specifically, some of them are: Typical asymptotic inference is inappropriate, because sample sizes are too small.

Multiple genes observed in the biological samples are likely to be correlated due to underlying regulatory circuitry of all genes and genes in various pathways. Hence, an assumption of independencies across genes is usually inappropriate. In cases that this assumption is required of, one needs to carefully assess the impact on computation of P-values when this assumption is violated. In fact, testing up- or down-regulation of genes in the same pathways would be of special interest in scientific investigation.

The high dimensionality becomes a challenge to any regression analysis, if all genes are treated as covariates in models, since the number of covariates exceeds the number of independent samples. Hence, it is essential to design a strategy of building appropriate statistical models.

When gene expressions were treated as outcomes, one would naturally be interested in modeling the joint distribution of multivariate gene expression values. However, if any likelihood method is used, one would have to make an assumption about multivariate outcomes, which needs to be justified and checked carefully. It is important

to assess impact on the likelihood inference if assumed distribution for the likelihood-based methods is violated.

Processing thousands of genes can be time-consuming. Hence, it is utterly important to devise either relatively simple methods, or fast computing algorithms for complicated methodologies.

Statistical analysis also must acknowledge biological reality. For example, handling and processing of biological samples may induce significant heterogeneities across microarrays. If not accounted for, such heterogeneity would increase estimated variations and would hence decrease the power of discoveries.

Multiple comparisons are another thorny problem for genomic data analysis with calculation of many P-values, since unadjusted P-values are at best misleading.

Statistical geneticists have had some experience with the multiple comparison issue in the context of genome scan studies for mapping complex traits. In the early day of linkage analysis, Morton recognized this problem of declaring a positive linkage with multiple genetic markers from human chromosomes (6), and proposed to use a lod score of 3 (or equivalently, log likelihood ratio=6.91) as the minimum threshold to declare the statistical significance. More recently, Lander and Kruglyak (1998) suggested raising this criterion to 3.6 to account for the fact that many more genetic markers are now being used in mapping studies (5). Their proposal is gaining an acceptance after much discussions and debates (8). The main rationale for increasing the threshold is to control the genomewide significance (the genomewide false discovery rate) at the 5% level. Indeed, this genomewide significance level is intrinsically connected to the statistical significance after adjusting for multiple comparisons (9). Lessons from mapping studies shall guide us in developing appropriate adjustment for multiple comparisons.

We are now in the Post Genomic era with the human genome map in hand (maps for many other organisms have been completed or will have been completed in the near future), and with ever-advancing biotechnologies. As statistically inclined researchers, we need to learn biology and biotechnology, and to appreciate that measurements in biology are frequently on a relative scale and have many sources of variation. We must also learn to be more effective in presenting critical statistical issues and solutions to our collaborators. In addition to devising complex models, we should develop robust and efficient methods to answer genomic questions, with valid assessment of P-values. P-values, quantifying the statistical significance, are one of the few available measurements for scientists to communicate effectively with each other, with the public, with politicians as well as with governmental regulatory agencies. It is simple, intuitive, scale-free and, most importantly, reproducible. Few substitutes for P-values exist!

References

1. Botstein D, White RL, Skolnick M, Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. American Journal of Human Genetics 32[3], 314-31. 1980.

2. Brent R. 2000. Genomic Biology. *Cell* 100:169-83

3. Collins F, Galas D. A new five-year plan for the U.S. Human Genome Project. Science 262, 43-6. 1993.

4. International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. *Nature* 409:860-921

5. Lander ES, Kruglyak L. 1996. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genetics* 11:241-7

6. Morton NE. 1955. Sequential tests for the detection of linkage. *American Journal of Human Genetics* 7: 277-318:

7. Venter JC, Adams MD, Myers EW, Li PW, et al. 2001. The Sequence of the Human Genome. *Science* 291:1304-51

8. Witte JS, Elston RC, Schork N. 1996. RE: Genetic dissection of complex traits. *Nature Genetics* 12:355-6

9. Zhao LP, Prentice RL, Shen FM, Hsu L. 1999. On the assessment of statistical significance in disease gene discovery. *American Journal of Human Genetics* 64:1739-53

Table 1. Frequencies of entries in PubMed with various key word/phrases in combination with "Statistical Significance" (total entries= 13,724 by "Statistical Significance" alone) and also by themselves used as denominators.

Applied Research	Frequency / Denominator	Basic Science Research	Frequency / Denominator	Genomic Research	Frequency / Denominator
Human	11,310 / 7,484,478	Biology	70 / 118,522	Genomics	1 / 2,226
Animal	2229 / 3,222,742	Molecular Biology	0 / 22,374	Functional Genomics	0 / 635
Biomarker	616 / 220,395	Cellular Biology	0 / 4,784	Proteomics	1 / 692
Population	1290 / 337,480	Biochemistry	321 / 280,236	SNP	4 / 2,687
Epidemiology	2351 / 659,132	Toxicology	20 / 11,501	GeneChip	1 / 48
Medical	2623 / 1,200,041			Microarray	6 / 1,114
Clinical	3521 / 1,096,438			Array	28 / 19,389
Pathology	1996 / 1,112,776				

!!! Controversial !!! Statistical Issue

Missing Data in Clinical Trials

What is Missing in the Missing Data Controversy

Guest Editorial: Robert T. O'Neill, Ph.D.

Robert T. O'Neill, Ph.D.
Office of Biostatistics
Center for Drug Evaluation and Research
US Food and Drug Administration

Clinical trials are the primary experimental mechanism used to establish the efficacy and safety of a new medical product or to compare the relative benefits of two or more products. Two critical components of the infrastructure of the modern clinical trials are the prospective statistical planning of the study design and the prospective planning for the subsequent implementation of a variety of statistical data analyses that produce the quantitative estimates and uncertainties associated with estimates for the chosen study design and the hypotheses tested. Robustness analyses are usually needed, either because of events, outcomes and unplanned or unexpected situations that occur during the conduct of the trial which lead to missing information on subjects randomized or when assumptions of the statistical analysis plans are not met. These analyses assess the impact of bias associated with the estimate of the treatment effect and the relationship of these biases to type 1 error, type 2 error, and statistical strength of the evidence. Missing data is often at the core of these analyses. In this issue of the Bulletin, several authors comment on various aspects of the missing data problem.

Missing data in the form of unplanned, unintended, and unobserved patient measurements during the conduct of a trial that occur prior to planned completion of the trial is more common than we would like because human experimental studies depend upon compliance, cooperation and contingencies of everyday life of the subjects participating in the trial. Statistical planning relies on the assumption that observed outcomes of all randomized subjects will be achieved. Missing data is a major challenge to the success and the validity of the conclusion of every trial, and is a complication that should be avoided when possible and, at the very least, should be anticipated and planned for in advance of trial conduct. When it does occur, the solutions are many but the consensus on the best approach is missing. Usually an in depth discussion in the protocol, or in the protocol analysis plan of the expected missing data situations, how it will be dealt with, and the interpretations may depend upon the approach used is missing.

The issues involve whether the unobserved measurements are an outcome in their own right, or whether they are a nuisance that can be overcome by statistical strategies that can replace in some sense what was not observed. Inherently, the problem is different for a variety of patterns of incomplete patient level response information, repeated measures designs, time dependent covariate information, and survival, time to event and absorbing event states for which no future data would be measurable anyway. While missing data statistical strategies seem to be successful in areas that face missing data, such as survey samples, there is debate about whether and how the design or analysis of clinical trials can benefit from these strategies.

It is my opinion, that for clinical trials, one should generally begin from the perspective that missing data on any patient randomized should be presumed to be informative unless demonstrated otherwise. Without getting into current definitions of types of informative, non-informative missing data or ignorable or non-ignorable types, it is worth considering in advance how the missing data is going to be treated in the analysis. Aside from defining in advance a missing pattern patient level response vector that is analyzed as a study outcome, most of the solutions seem to rely on selecting models, either to impute what is missing or to adjust estimates after the fact for the data that is censored (missing). Which model is best, what is a metric for judging best, and how should the bias be quantified and adjusted for, all seem to remain open questions. These types of analyses are not routine in current practice, except for the most simple strategy that impute outcomes by using the last observation carried forward strategy. Under certain circumstances, this latter approach may be conservative and simple enough to provide an answer that is robust. In other circumstances, any imputation that does not simultaneously incorporate variability and location, and bias can be considered less than desirable, but even with their use we are left asking the question of what is the best metric to evaluate the resulting tests of hypotheses and estimates of treatment effect. Results that depend upon extensive analyses and assumptions, might need substantial evaluation themselves by independent referees or reviewers who are not tied to the models and approaches used by the analysts.

Should one routinely pretest all missing data problems for whether they reflect random missingness, informative missingness, or a combination of the two? Is there such a pretest that distinguishes among these? Whatever is decided, it seems that virtually all solutions to the missing data problem involve appealing to models and/or regression strategies that are used to impute what was not observed. The alternatives seem numerous but how does one decide on the validity of the conclusion?

Prospective simulation to evaluating the impact of missing data, dropouts, multiple competing risk outcomes, and multiple censoring distributions at the protocol planning stage seems a worthwhile effort, if only to make apparent the logic of what will be done at the data analysis stage, and to raise the consciousness of clinical trial planners as to what could be expected under a variety of situations, what analyses would be done in response to those scenarios, and what are the limitations to inference. Patient withdrawal from treatment should not mean that future measurements while in the trial are not possible. Data missing on a subject because no effort was made to collect information on that subject after treatment withdrawal is a practice that is missing the reason for doing the trial in the first place. Patient withdrawals are outcomes in their own right, but not necessarily absorbing events. Conditional on a patient withdrawal, subsequent subject outcomes and measurements can be of use. Advance planning about this issue is missing and never considered until after the fact when the data is in hand, and when it may be too late to use any method that satisfactorily addresses the issue.

Missing Data: A Damaging (And Possibly Catastrophic Disorder) For Statistical Planning

Gary G. Koch, Ph.D.
Department of Biostatistics, CB#7420
School of Public Health
University of North Carolina, Chapel Hill, NC, USA, 27599-7420

The primary objective of statistical planning for any inferential study is the production of a data structure for which the results from analyses are reasonably convincing. Conscientious statistical planning for this objective is particularly important for randomized clinical trials for the comparison of two or more treatments with respect to one or more criteria for the health or well-being of the participating patients. The randomized clinical trial has both its clinical and statistical plans specifically expressed in its protocol. Vital statistical components of the protocol include the statistical expression of the primary hypotheses, the planned methods for statistical analyses to address the hypotheses (with appropriate attention to such issues as multiple comparisons, adjustments for covariates, nonparametric or regression model specification), and the justification for the sample size (in terms of provision of sufficient power for the clear evaluation of the primary hypotheses with the planned methods for analyses). In the ideal (but very unusual) situation where all patients in a randomized clinical trial completed its specified time period for treatment and follow-up with strict compliance to its protocol and its schedule and scope for data collection, the results from the planned analyses for the primary hypotheses would be reasonably convincing [1].

Unfortunately, nearly all randomized clinical trials have patients who do not comply with the schedule and scope for data collection; and among such patients, those who do not complete the time period for treatment and follow-up require careful attention because their subsequent status with respect to their continuation of the protocol to completion is entirely unknown. The unknown nature of the missing data after the withdrawal of patients who do not complete a study (as well as missing data at times prior to withdrawal) undermine the inferential structure of a study. Moreover, the corresponding damage can catastrophically imply that the observed data from a study are not interpretable when the prevalence of missing data is excessive. The way in which missing data damages the inferential structure of a study is through their impact on the applicability of the planned analyses for a study.

The crucial issue for the patients with missing data is that they do not provide sufficient information for inclusion in analyses in the originally planned manner, and so the undermining damages from missing data are the unplanned modifications which it causes for analyses. Moreover, whether such modifications are appropriate or not is unknown because of the unknown nature of the missing data that they address. For this basic reason, there is no "clearly correct" method for managing missing data in analyses for which the original statistical planning is for a data structure with no missing data. Thus, a randomized clinical trial with missing data must have analyses by two or more methods which manage missing data according to arguably reasonable and planned principles; and the results from such analyses must agree with one another in order to be convincing in a sense that is robust to missing data [1]. Some of the methods with planned principles for missing data include replacement of missing data by a most favorable value, replacement by a least favorable value, replacement by the most recent previous value, replacement by an estimate from a "reasonable statistical model" and maintenance as missing [2]. These methods have planned principles because they specify how missing data are managed, but the missing data that have their implementation are unplanned and unknown, and so the correctness of these methods is unknown. As stated previously, the results from a randomized clinical trial need to be robust to missing data in order to be convincing, but such robustness may need to account for "punitive methods" such as replacement of missing values for the test treatment with "bad" values and replacement of those for the control treatment with "good" values; and so robustness to a convincing extent can be reasonably likely only when the prevalence of missing data is minimal. Thus, the most appropriate statistical strategy for managing missing data is to minimize its prevalence through sufficiently well planned study designs and procedures for data collation [1].

A reasonably straightforward strategy for strengthening the robustness of a study design to missing data is to have the protocol (or the informed consent) specify continuation of data collection regardless of whether a patient stops using the randomly assigned treatment (or has a major protocol violation). A very important property of this supplementary data is that it is not missing. However, it has the limitation of not being in accordance with the protocol, and so some careful planning is necessary for its role. In some situations, the supplementary data after a patient discontinues their randomly assigned treatment can have direct inclusion in a primary response variable on the basis of its clinical relevance (e.g., survival, disease progression). An alternative role for this supplementary data is its potential capability for providing indirect support to an a priori planned principle for managing missing data (e.g., replacement by the most recent previous value, replacement by the worst possible value, etc.). This point is noteworthy because the supplementary data can sometimes sufficiently clarify the robustness of the results from the analyses with the planned principles for the management of missing data that other less powerful assessments of robustness are not logically necessary (e.g., punitive principles such as replacement of missing values for the control group with "good" values and replacement of missing values for the test treatment group with "bad" values).

The supplementary data after patients discontinue their randomly assigned treatments can alternatively have inclusion in nonparametric assessments according to Mann-Whitney principles [3,4]. In such assessments, each patient has classification of their overall response (or of multiple responses for two or more criteria and /or two or more time points) as better, tied or worse relative to each other patient (with the pairwise structure usually involving each patient with a test treatment versus each patient with a control treatment). This principle can address supplementary data by assigning "better" to a patient who has "better on treatment data" than another patient's supplementary data, but assigning "tied" to a patient who has " poorer on treatment data" than another patient's supplementary data (with the justification for "tied" being the longer use of the randomly assigned treatment for the former patient). These pairwise assessments can then be combined in nonparametric rank tests for the comparison between the test and control treatments, or in rank measures of association [4] for the probability that a randomly identified patient with test

treatment has better response than a randomly identified patient with control treatment. In such analyses, a noteworthy advantage for nonparametric methods is that their use is possible without possibly unrealistic assumptions concerning the underlying distributions of the data. This point applies as well to methods which involve replacement of missing values according to planned principles since the replaced values do not usually satisfy the assumptions concerning underlying distributions that pertain to the values that are missing. For nonparametric methods, the relevant assumptions for statistical tests are randomization in the study design and validity of the data being analyzed [3]. For rank measures of association, the first of these assumptions needs to be strengthened to the patients in each treatment group being comparable to a simple random sample from a corresponding population [4]. These assumptions are realistic for well conducted randomized clinical trials with no missing data as well as to those for which the observed data structure and the planned methods of analyses provide robustness to their realized extent of missing data.

Another strategy for reducing missing data through the design of a randomized clinical trial is the use of a two period crossover design (with the first period as the primary data structure for assessing hypotheses). The purpose of this design would be to enhance completion of the first period by patients and correspondingly to have better compliance with the protocol as the prerequisite for access to the treatment during the second period [1]. Since the first period is the primary basis for statistical inference in this type of study, issues concerning carryover effects (or treatment x period interaction) are not critical, and neither is discontinuation of patients after completion of the first period, particularly if they no longer satisfy the entry criteria for the second period (perhaps as a consequence of their favorable outcome for the first period). For this use of a crossover design, the main role of the second period is supplementary; and a design of particular interest has three sequence groups as follows: test:test, test:control, control:test [1]. The second period of this design enables evaluation for whether a second period of test treatment provides better outcomes than withdrawal of test treatment after the first period, and whether outcomes at the end of the second period for patients who received control treatment during the first period are comparable to those for patients who received test treatment during the first period only or both periods. Regardless of such roles for the second period, the dominant purpose for this use of a crossover design is the minimization of missing data during the inferentially primary first period of the study.

REFERENCES

[1] Koch, G.G., Davis, S.M., Anderson, R.L. Methodological advances and plans for improving regulatory success for confirmatory studies. Statistics in Medicine, 1998; 17; 1675-1690.

[2] Little, R.J.A. and Rubin, D.B. Statistical Analysis with Missing Data, Wiley, New York, 1987.

[3] Koch, G.G., Tangen, C.M., Jung, J.-W. and Amara, I.A. Issues for covariance analysis of dichotomous and ordered categorical data from randomized clinical trials and nonparametric strategies for addressing them. Statistics in Medicine, 1998; 17; 1863-1892.

[4] Jung, J.W. and Koch, G.G. A linear model method for rank measures of association from longitudinal studies with fixed conditions (visits) for data collection and more than two groups. Journal of Biopharmaceutical Statistics, 1998; 8(2); 299-316.

Issues in Handling Missing Data Caused by Dropouts in Clinical Trials and Multiple Imputation

Weichung J. Shih
University of Medicine and Dentistry of New Jersey School of Public Health

I guess missing data was not exactly what the Bible tried to inspire, but Saint Paul did write to Christians of the Corinth Church in 55 AD and encouraged them to ' ... *fix our eyes not on what is seen, but on what is unseen.*' (Second Corinthians 4:18). In this brief note I'd like to share with my fellow ICSA Bulletin readers my viewpoints when I fix my eyes on data sets from clinical trials that contain 'what is unseen'.

1. Notation Expressed in general terms, let Y denote the complete-data matrix of the outcome variable, which follows the distribution $[Y] = f(Y)$, where, for convenience, the $[\bullet]$ denotes the distribution of the designated random variable inside the bracket without referring to a specific parametric model. When missing data occur, Y is the '*hypothetically complete data*' and can be written as $Y = (Y_{obs}, Y_{mis})$, where Y_{obs} is the observed part of Y (termed the '*incomplete data*') and Y_{mis} is the missing part of Y. Furthermore, let M denote the missing-data indicator matrix corresponding to $Y = (Y_{obs}, Y_{mis})$. M is the so-called '*missing data pattern*', which is also a random variable, and its distribution in relation to Y, $[M| Y]$ is called the '*missing data process or mechanism*'. The '*data set*' actually is (Y_{obs}, M), not Y_{obs} alone. In addition, the underlying reason for the missing data or dropouts is also an important part of the information which should be considered in the analysis, and there may be predefined baseline covariates, X, which are relevant to the data as well. In our notation, $X_1 \subset X$ will be the treatment group indicator covariate.

2. Objective of a Data Analysis As in any data analysis, the first consideration is the objective of the analysis, which can be formed by the question or questions the analysis intends to address. In the situation where the complete data become hypothetical, this can be an issue as described in the following.

Since the missing data pattern M is a part of the data, we need to consider the joint distribution of Y and M, $[Y, M]$, which can be written in two different ways: $[Y, M] = [Y]\,[M \mid Y]$, or $[Y, M] = [M]\,[Y \mid M]$. Adding predefined baseline covariates in the above formulation and attaching the specific parameters associated with the various distributions, we then have:

$$[Y, M \mid X] = [Y \mid X; \theta]\,[M \mid Y, X; \Psi] \quad (1)$$

or

$$[Y, M \mid X] = [M \mid X; \eta]\,[Y \mid M, X; \phi]. \quad (2)$$

When there is no missing data, there is only one possible parameter for inference, that is, θ. However, when there is missing data, the possible parameters of interest for inference include θ, (θ, Ψ), or (η, ϕ). In other words, we need to clarify the question and define the objective of the analysis as whether we are addressing:

comparison of the treatment groups in terms of θ or (θ, Ψ) in the hypothetical complete data when all patients follow the full therapy course, or

comparison of the treatment groups with the incomplete data in the presence of dropouts in terms of (η, ϕ).

It is possible, in light of the ICH-E9 guideline, that both questions have to be addressed in the data analyses, but most missing data literature focus on the question regarding θ only,

including the method of multiple imputation (MI). Shih and Quan (1997, 1999) argued that in clinical trials, (η, ϕ) or part of (η, ϕ) may be more pertinent to medical questions in the presence of dropouts.

3. Multiple Imputation or Partial Imputation: MI is a useful and convenient method. It is useful since it imputes data 'properly' (i.e., providing both within and between imputation variations). It is convenient since one may use the usual complete data method for each set of imputed data. However, users need to understand that it, like any other method, has model assumptions too. Not only it assumes missing data are missing at random (MAR) in Rubin's sense (1976), the fill-in data has to be based on a certain regression model. The MI method is not a 'one size fits all' solution to all missing data problem, as some may suggest. Wang (1999) showed that MI can introduce substantial bias in certain non-ignorable missing data or mis-specified regression MAR model situations. In fact, no method can do the 'one-size fitting all' job in the missing data business. (It is a real business now if we read AMSTAT advertisements.) If the missing data mechanism cannot be specified correctly, as it often is the situation in practice, Wei and Shih (2001) introduced a partial imputation (PI) strategy. They argued that we should not impute all the missing values when we cannot be sure that our imputation model is correct. Instead, they suggest we only need to impute 'just right' to balance the dropout pattern to reduce the potential bias when comparing treatment groups. As a new idea, PI remains to be further investigated.

4. Conclusion: The issue of missing data due to dropouts in clinical trials is a current research topic that is still under development in the statistical literature. As commented in the ICH-E9 guideline, 'no universally applicable methods of handling missing values can be recommended.' The issue of handling missing data has an intrinsic difficulty: It requires a large proportion of missing data to investigate a method. On the other hand, a large proportion of missing data would make a clinical study less credible. The best available advice is to minimize the chance of dropouts at the design stage and during trial monitoring. If a large amount of dropouts is anticipated, then perhaps the study duration should be shortened or the medical procedure which likely causes patient's withdrawal needs to be altered. Consideration may also be given to define an endpoint (event), instead of a measurement value, as the primary response variable which can be followed up despite of patient's withdrawal from the study. In an analysis, one should be clear about the question or objective of the analysis with missing data, and conduct sensitivity analysis with possible models of the missing data.

I began with a Bible verse, and I shall end so. Deuteronomy 29:29 said, '*The concealed things belong to the Lord our God, but the things revealed belong to us and our children....*' I humbly submit.

References

Rubin, D.B. (1976). "Inference and missing data", *Biometrika*, 63:581-592.

Shih, W.J. and Quan, H. (1997) "Testing for Treatment Difference With Dropouts Present in Clinical Trials -- A Composite Approach" *Statistics in Medicine*, 16:1225-1239.

Shih, W.J. and Quan, H. (2001) "On the Composite Approach to Dropouts in Clinical Trials" *Statistica Sinica*, 11:53-62.

Wang J. (1999) "The Application of Multiple Imputation Using a Commercial Software in Clinical Trials with Dropouts", American Statistical Association 1999 Joint Meetings, Proceedings of Biomectrics Section, 53-60.

Wei, L. and Shih, W.J. (2001) "Partial Imputation Approach to Analysis of Repeated Measurements With Dependent Dropouts" *Statistics in Medicine*, 20:1197-1214.

Some Issues with Handling Dropouts in Analysis of Clinical Trials*

H.M. James Hung, Ph.D.
Division of Biometrics I
Office of Biostatistics
Center for Drug Evaluation and Research
US Food and Drug Administration

*The views expressed in this article do not necessarily represent those of the U.S. Food and Drug Administration

Needless to say, handling missing data is one of the difficult tasks in statistical analysis. Literature is abundant on this subject, such as, Little and Rubin (1987), Little (1995), Laird (1988), etc. Statistical methods for managing missing data often involve extensive mathematical work and assumptions. The statistical inference is likelihood-based or Bayesian. There is little doubt that for decades the methodological advances in this area have contributed much to biomedical research and helped to explore how the management of dropouts in statistical analyses affects the interpretation of the studies.

Despite the methodological advances, scientists working in the clinical trial area continue to struggle in figuring out how to best handle dropouts in statistical analysis. The difficulty is more transparent in analysis of the so-called 'confirmatory' trials. In fact, as often is the case, the dropouts are assigned a value for analysis by some 'natural' simple-minded imputation approaches, such as, last value carried forward, worst value, worst rank, etc. The 'natural' approaches are by all means highly questionable scientifically and often render the results that can be interpreted from one extreme to the other and hence are little informative.

The foundation of statistical inference for randomized clinical trials is, in principle, based on the randomization distribution. The normal theory or the likelihood theory often employed in the analysis is intended to generate an approximation to the randomization-based inference. Under the randomized-based framework, there is no probabilistic structure available for describing how a patient's dropping out of the study is related to the outcomes, missing or not. This is sad of course but the framework of inference is what it is. Should one then change the framework of inference to likelihood-based inference? What is the price to be paid for the potentially big benefit by doing so?

Consider a randomized clinical trial for comparing the effectiveness between an experimental therapy and a control. The trial is planned to seek the comparison at the time after a period of some length (8 weeks, say) passes. All randomized patients are to be included in the comparative analysis. Almost surely, some patients will drop out of the study for a variety of reasons, such as, lack of effect, undesirable or intolerable side effects, etc. At the end of the trial, the treatments are compared using the last available or last visit value from each patient. If the dropout patterns in the clinical trial mimic the patterns in real clinical practice, then this last value analysis is useful, at least, for global assessment of the performance of the experimental treatment. This assessment involves a different timing of observation from patient to patient and hence does not accurately describe the treatment difference at the specific end time. The last value analysis confounds with time effect if any and may confound with other factors like reasons of dropouts; therefore, like any analysis it may be difficult to tease out the efficacy of the experimental treatment. However, an important point is that the last value analysis is conceptually very different from the last value carried forward analysis, though they can be made equivalent mathematically. The last value carried forward analysis aims at the performance of the experimental treatment at some specific fixed time point, whereas the last value analysis is meant to give a global assessment under the real clinical practice. So the interpretations of the results of the two analyses are different.

It is certainly a serious concern that the patient's dropping out of the study is influenced by the investigators in the trial in some biased way and so the dropout patterns do not mimic the real clinical practice. When this is suspected, a conservative way for assessing the treatment difference will need to be sought. The question is how under the randomization-based inference framework? None of the 'natural' approaches are satisfactory. But the imputation methods prevailing under the framework of likelihood-based or Bayesian inference offers little help for one obvious reason given above. That is, under the randomization-based inference, there is no probabilistic structure available for describing how a patient's dropping out of the study is related to the outcomes, missing or not. Placing mathematical structures that is necessary under the likelihood or Bayesian inference may help to explore the impact of dropouts but not without cost. What is the cost? At least, the cost could come from the potential bias due to unverifiable model misspecification that may have a dramatic impact on the effectiveness estimate and the standard error. More importantly, will the cost impede our ability to make a 'credible' conclusion for the clinical hypothesis at hand as we would have been able to have there been no dropout?

The aim of the clinical trial may well be the comparison of the effectiveness profile over the entire or some specific period of the trial. Perhaps it should be. The clinical question or hypothesis can be made more sensible with an intelligent help from statisticians. A more sensible efficacy parameter can be rate of change in pain scale over time, average reduction of blood pressure over the period at steady state, global assessment of degradation of the effect after a certain period of long-term use, etc. Like the last value analysis, the assessment on all the measurements in the relevant period of the study can be useful for global evaluation of the treatment differences, taking into consideration of dropouts under the scenario mimicking the real clinical practice. In analysis of such response profiles, it obviously makes no sense to carry the last value forward to impute all the future missing data and thus there should be no confusion like that between the last value analysis and the last value carried forward analysis. Now the bigger question is how to do sensitivity analysis (sensitivity to the treatment of missing data). By the way, what is the definition of the within-subject correlation in measurements between visits under the randomization-based inference?

Despite the questions described above, likelihood based methods may be very useful for analysis of randomized clinical trials if the resulting inference is carefully made together with consideration of the randomization process. At least, the models for missing data mechanisms may offer valuable insights about the robustness of statistical analyses against the management of dropouts. However, the opportunity of statistical modeling for dropouts will be limited by the usual concern with a high rise in false positive errors as a result of heavy multiple testing. Thus, the more urgent task is seeking statistical analysis methods (either likelihood or non-likelihood based) with a stable false positive (or type I) error regardless of how the dropouts are treated, rather than finding the most proper stochastic model for dropouts and then finding an 'optimal' analysis method under this dropout model. Another approach is working through the trial design to understand the probable profiles of dropouts using some kind of clinical trial simulators and then minimize the occurrence of dropout when conducting the clinical trial. However, even if we are able to achieve that, in all probability the clinical trial may not render the results that are relevant to the real clinical practice; for instance, it may overestimate the beneficial treatment effect and underestimate the chance of dropout occurring in real clinical practice.

Selected Reference

Laird, N.M. (1988), "Missing data in longitudinal studies", *Statistics in Medicine* 7, 305-315.

Little, RJA (1995), "Modeling the drop-out mechanism in repeated-measures studies", *Journal of the American Statistical Association* 90, 1112-1121.

Little, RJ, and Rubin, DB (1987). Statistical Analysis with Missing Data. Wiley: New York.

MISSING DATA IN CLINICAL TRIALS: WHAT TO RECOMMEND AN ACADEMIC CONSULTANT'S VIEW

Ralph B. D'Agostino, Sr.
Boston University, Boston, MA 02215
January 8, 2002

Clinical trials always have some missing data. If the data are missing completely at random and good balance remains between the treatments (i.e., the original randomization is not impaired greatly) then the missing data can be ignored. However, if the missingness relates to the treatments or to the outcomes, then analyses of treatment difference using only the observed data can produce biased estimates of treatment effects and inflated Type I errors. In response to this, statisticians have generated an array of imputation methods. Some non-statisticians even believe there is no longer a problem, since the statisticians now have many methods to deal with missing data. Statisticians may be encouraging sloppiness in clinical trails since they spend so much time on developing methods to deal with the problem, rather than trying to stop the problem form happening. Faced with these realities how is a consultant to advise sponsors or government agencies concerning missing data. There are a number of recommendations I believe can and should be made.

First, it is important to design a clinical trial so that the possibility of missing data is minimized. It is a seriously wrong attitude to wait until the damage is done and the try to fix things with some "overly sophisticated, assumption driven" imputation. There are many ways this advice can be implemented. For example, design the study so that the participants have motivation to return for visits and supply necessary information. Think out carefully how data will be obtained on outcome variables, even when the participant is not thoroughly involved. For treatments that might produce dropouts because of side effects, plan strategies that will keep subjects involved. If a placebo is used in the study, consider rescue medication or drop-in to the other treatment. Each study will have its own special features. The point is that we often know the design features that produce missing data. We should plan our studies so that they are minimized.

Second, it is important to execute a clinical trial so as to minimize missing data. Further, in the execution of a trial, this will require understanding the mechanisms that lead to missing data. Strategies to achieve these goals need to be developed and implemented. This recommendation is similar to the above, but now extends it into the operations of the trial and also attempts to incorporate mechanisms.

There are many examples we can give here. Subjects moving away from the study site area or subjects losing interest may results in data "missing completely at random." Increase in sample size may be sufficient to deal with this. Human errors and laboratory problems that result in missing data can be addressed by careful study execution and monitoring. Often missing lab data would not exist if back up blood samples were routinely taken. Missing clinic data often can be limited by rapidly applied quality control procedures. Missing data due to treatments (missing at random data) or outcomes may not be easy to prevent. So identifying how they will occur, monitoring features of the study where their effect will be manifest, and taking correct measures are essential. Strategies to minimize missing data are imperative.

Third, no matter what is done above, the clinical trial will have missing data and we will have to do analysis. Imputation methods must be selected. Begin the process with identifying imputation methods that are simple, robust, and as much as possible, model and assumption free. Make sure these are appropriate to the data of the study. I favor methods such a Wittes et al (1989), where missing dichotomous the observed proportion of events in the other treatment replaces outcomes in one treatment, rather than a more sophisticated imputation such as might result from random effects modeling (Heyting et al, 1992, Little and Rubin, 1987) or multiple imputation methods (Rubin, 1987). I believe it is always important to be able to explain what was done to the nonstatisticians. The Wittes et al (1989) method has an intuitive appeal. Other procedures, for example, such as post stratification methods employing propensity scores (Lee, 2001) and pattern mixture modeling approaches (Shih and Quan (1997) and Miller et al. (2001)) may also be. The important point here is to select the methods mainly because of their intuitive appeal and robustness.

Fourth, identify conservative methods and justify their use or non-use in the particular setting. The Last Observation Carried Forward (LOCF) method is, in my experience, often a useful method to apply. It may in fact be the method selected in the above third step. In an analgesic study where the condition under investigation is improving over time and "dropouts" may be mainly associated with taking rescue medication, the LOCF may work very well. In situations dealing with psychiatric medications it may not (Lavori, 1992). In this context the array of methods considered could also include modified LOCF such as developed by Wei and Shih (2001).

Fifth, consider how sophisticated and model dependent we can be. Here the random effects methods of Heyting et al. (1992) and Laird (1988) and the multiple imputation models of Rubin (1987) should be considered. If justifiable, these are wonderful methods.

Sixth, and here is a hard one, select an imputation method from steps three to five above that is as simple as possible, as model free and robust as possible, and that takes into account the missing data mechanisms that we identified in previous considerations. Software exists or can, with reasonable effort, be generated for many of the methods identified above and so selection need not be software-existing dependent. I usually encourage a method identified from step three as the final method.

While it is possible to do analyses employing a number of imputations schemes, I advise against it. It is easy to fall into the mode of trying everything and selecting the one that matches best our desires and biases. The one selected should have a priori justification.

Seventh, perform a sensitivity analysis with the imputation method selected in step six. Some would cloak this in the setting of a Bayesian analysis. The important point is to investigate how inferences change with changes to the selected imputation scheme.

Even after all the above is done, there may well exist some missing cases where we can obtain further information. For example, in a large simple cardiovascular trial, there will be dropouts on whom we will not be able to obtain information as to their cardiovascular events. However, we may be able to find out if they are alive. While we may not be able to

incorporate this data formally into the above analyses, it often helps us interpret the analyses.

CLOSING COMMENTS

A clinical trial without missing data is probably impossible in any realistic setting. Our first reaction to it should be to improve what we are doing so that the missing data is reduced in the future, and not to immediately think we can fix it with imputation methods. When faced with the need to do something with the missing data, we should try to deal with it with methods that are simple, explainable, justifiable and robust. We should be aware of the effect of conservative methods applied to the data. We should not avoid pushing the data and our thinking to decide when assumption-heavy techniques such as random effects and multiple imputation methods are appropriate. In all cases, sensitivity analyses should be performed to understand the stability of our results. Lastly, when consulting on missing data issues, I always return to the beginning. We must avoid sloppiness in clinical trials by leading people to believe we can fix the problem by use of clever statistical methods. We must design and execute clinical trails that have minimum missing data.

REFERENCES

Heyting A, Tolboom JT, Essers JGA. Statistical handling of dropouts in longitudinal clinical trial. Statistics in Medicine 1992: 11:2043-2061.

Laird NM. Missing data in longitudinal studies. Statistics in Medicine 1988; 7:305-315.

Lavori PW.Clinical trials in psychiatry: should protocol deviation censor patient data? Neuropsychopharmacology 1992; 6:61-63.

Little RJA, Rubin DB. Statistical Analysis with Missing Data. 1987. Wiley, New York

Lee YJ. Post-stratification based direct adjustment approach to a missing data problem in clinical trials. Journal of Statistical Planning and Inference 2001; 96:247-262.

Miller ME, Morgan TM, Espeland MA, Emerson SC. Group comparisons involving missing data in clinical trials: a comparison of estimates and power (size) for some simple approaches. Statistics in Medicine 2001; 20:2383-2397.

Rubin DB. Multiple Imputation for Nonresponse in Surveys. 1987 Wiley, New York.

Shih WJ, Quan H. Testing for treatment differences with dropouts present in clinical trials-a composite approach. Statistics in Medicine 1997; 16:1225-1239.

Wei L, Shih WJ. Partial imputation approach to analysis of repeated measurements with dependent drop-out Statistics in Medicine 20;1197-1214.

Wittes J, Lakatos E. Probstfeld J. Surrogate endpoints in clinical trials: cardiovascular diseases. Statistics in Medicine 1989; 8:415-425.

Some Upcoming Statistical Meetings

ProbaStat 2002
February 04 – 08, 2002, Smolenice Castle, Slovak Republic

International Biometric Society Eastern North American Region 2002 Spring Meeting with IMS and Sections of ASA
March 17 – 20, 2002, Virginia, USA

UConn Department of Statistics: A Day of Celebration and Dedication
April 06 – 06, 2002, Storrs, Connecticut, USA

Interface '02, the 34th Symposium on the Interface of Computing Science&Statistics
April 17 - 20, 2002, Montreal, Quebec, Canada

Conference on Data Mining, Cleveland, Ohio
May 06 – 06, 2002, Cleveland, Ohio, USA

Society for Clinical Trials Annual Meeting
May 12-15, 2002 - Arlington, VA, USA

ICSA 2002 Applied Statistics Symposium at Greater Philadelphia, Pennsylvania
June 06 – 08, 2002, Philadelphia, Pennsylvania, USA

International Conference on Computer-Based Testing and the Internet
June 12 – 15, 2002, Winchester, United Kingdom

2002 Taipei Intl' Statistical Symposium&Bernoulli Society EAPR Conference
July 07 – 10, 2002, Taipei, Taiwan, R.O.C.

Current Advances and Trends in Nonparametric Statistics,Co-sponsored by IMS
July 15-19, 2002, Crete, Greece
http://www.stat.psu.edu/~npconf/

XXIth International Biometric Conference IBC2002
July 21 – 26, 2002, Freiburg, Germany

Third International Conference on Multiple Comparisons Procedures
August 04 – 07, 2002, Bethesda, Maryland, USA

2002 Joint Statistical Meetings
August 11 – 15, 2002, New York, USA

ICSA 2003 Applied Statistics Symposium
June 22 - 24, 2003, University of San Diego, San Diego, California, USA

ANNOUNCEMENT (12/15/2001)

ICSA 2002 APPLIED STATISTICS SYMPOSIUM
JUNE 6-8, 2002

Theme: The Leading Edge of Statistics in Health Sciences
Location: Doubletree Guest Suites, Plymouth Meeting, Greater Philadelphia

KEYNOTE SPEAKERS (June 7, Friday, 8:45 AM – 10:15 AM):

- *Dr. Robert O'Neill*, Director, Office of Biostatistics, CDER, FDA
- *Dr. George Williams*, Vice President of Biostatistics and Programming, Pharmaceutical Research Institute, Bristol-Myers Squibb Co., and Vice President of ASA

DINNER SPEAKER (Chinese Banquet on June 7, Friday, 7:00 PM – 11:00 PM):

- *Judge Ida K. Chen*, Court of Common Pleas, First Judicial District of Pennsylvania.

PLENARY SESSION SPEAKER (June 8, Saturday, 8:30 AM – 10:00 AM):

- *Dr. Gordon Lan*, Pfizer Inc.
  Topic: Adaptive Design and Analysis

SHORT COURSES (June 6, Thursday): (See later pages for details)

Course Title	Instructor(s)	Time
1. Design Considerations for Positive Late Phase Confirmatory Trials	*Irving Hwang, Ph.D.* University of Medicine & Dentistry of NJ, Irving Consulting Group	9:00 AM –5:00 PM
2. Active Control Non-Inferiority Studies and Adaptive Analysis Methods in Clinical Trials	*Sue-Jane Wang, Ph.D. & James Hung, Ph.D.* CDER, FDA	9:00 AM –5:00 PM
3. Statistical Approaches in Pharmacogenomics	*Kim Zerba, Ph.D., Shu-Pang Huang,Ph.D. & Frank Shen, Ph.D.* Bristol-Myers Squibb	9:00 AM –5:00 PM
4. Advanced Log-linear Models for Categorical Data With GENMOD	*Daniel Zelterman, Ph.D.* Yale University	9:00 AM –5:00 PM

INVITED SESSIONS (June 7-8, Friday & Saturday):

Session Topic	Organizer	Speakers (not all complete)
1. Regulatory Issues in Planned Interim Analysis	*Gary Aras*	*Mohammed Huque*, CDER, FDA *Qing Lin*, R. W. Johnson PRI *Lu Cui*, Aventis Pharmaceuticals
2. Statistical Considerations in Evaluating Patient Reported Outcomes	*Joe Heyse*	*Joseph Heyse*, Merck Research Laboratories *Margaret Rothman*, Johnson & Johnson *Lisa Kammerman*, CDER, FDA *Dennis Revicki*, MEDTAP, International
3. Adaptive Two-stage Designs for Phase II Cancer Trials	*Joe Shih*	*Yu Shyr*, Vanderbilt University *Bee-Lian Chen*, Novartis Pharmaceuticals *Yong Lin*, UMDNJ
4. Design and Analysis of Active Control Non-Inferiority Trials	*Irving Hwang*	*Irving Hwang*, UMDNJ & Irving Consulting Group *George Y.H. Chi*, CDER, FDA *Antonella Maniero*, Bristol-Myers Squibb PRI *Yi Tsong*, CDER, FDA
5. Issues and Advancement in Pharmaceutical Statistics	*Agnes Hsiung*	*Jen-pei Liu*, NHRI, Taiwan *Chin-Fu Hsiao*, NHRI, Taiwan *Ivan Chan*, Merck Laboratories *Wayne C.S. Weng*, Schering Plough Pharmaceutical
6. Statistics Issues in Preclinical Research	*Shein Chow*	*Jun Shao* (chair), University of Wisconsin *Soomin Park*, Eli Lilly and Company *Lei Shen*, Ohio State University *Jinglin Zhong*, Merck Laboratories *Hansheng Wang*, StatPlus, Inc.
7. Statistical Application in Genomic Research	*Chi-Hse Teng*	*Frank Shen* (chair), Bristol-Myers Squibb *Laura Mitchell*, University of Pennsylvania *Kim E. Zerba*, Bristol-Myers Squibb *Chi-Hse Teng*, Pharmacia Corporation
8. Statistical Issues in Design and Analysis of Microarray Data	*Jeff Wu*	*Li-An Xu*, Bristol-Myers Squibb *Greg Dyson*, University of Michigan *Dan Holder*, Merck Laboratories
9. Statistical Data Mining in Early Drug Discovery	*Dhammika Amaratunga*	*Dhammika Amaratunga*, The R W Johnson PRI *Christelle Darstein*, Eli Lilly & Co *Nanxiang Ge*, Aventis Pharmaceuticals
10. Advancement of Computer Assisted Trial Design in Drug Development	*Peter Ouyang*	*Wenping Wang* and *Dan Weiner*, Pharsight Corp. *Wayne Ewy*, Pfizer Global Research & Development *Andrew T. Chow*, The R. W. Johnson PRI
11. Non-parametric methods in longitudinal analysis	*Jianhua Huang*	*Lijian Yang* (Chair), Michigan State University *Hulin Wu*, Harvard School of Public Health *Wenshen Guo*, University of Pennsylvania *Jianhua Huang*, University of Pennsylvania
12. Exact Conditional Inference for Categorical Data	*Darshong Hwang*	*Dar Shong Hwang*, BRSI and NHRI, Taiwan *Chyi-Hung Hsu*, Novartis Pharmaceuticals *James S. Lee*, Novartis Pharmaceuticals *Xin Wei Jia*, Janssen Research Foundation

13. Multiple Imputations of Missing Values	*Xiao Li Meng*	*Xiao Li Meng,* Harvard University *Joseph L. Schafer,* Penn State University *Nat Schenker,* CDC
14. Financial Econometrics	*Ruey Tsay*	*Rong Chen,* University of Illinois *Clara Vega,* University of Pennsylvania *Xu-Feng Niu,* Florida State University

CONTRIBUTED SESSION (June 7, Friday):
The committee invites you to submit statistical papers to be considered for presentation at the symposium. A small number of contributed papers will be accommodated. Please submit abstracts in text format by February 28, 2002, to Francis Hsuan, Temple University, at francish@vm.temple.edu for processing. The abstract should not exceed 200 words. Please include the name, affiliation, mailing address, telephone number, and e-mail address of the author.

STUDENT AWARDS and TRAVEL FELLOWSHIPS:
The committee encourages student members of ICSA to participate and present their research work at this annual meeting. Up to 3 travel award winners will be selected and notified by April 15, 2002 by the Awards Committee chaired by Prof. Weichung J. Shih. All winners will each receive a certificate, $400, and tuition for one short course of their choice. The manuscripts should be received and postmarked no later than February 28, 2002. Please see page 47 of the July 2001 issue of the ICSA Bulletin, or visit the ICSA website, http://www.icsa.org/, for detailed information regarding the submission.

HOTEL INFORMATION:
The symposium will be held at the Doubletree Guest Suites in Plymouth Meeting in Greater Philadelphia. To make a room reservation, please call 610-834-8300 or 800-222-8733. ICSA has arranged a special group rate, $119 per suite for one or two people, to be applied to reservations made on or before May 10, 2002. The hotel is extremely busy, so please make your reservation early. Reservations may be cancelled any time before 4 p.m. on the date of your arrival.

LOCAL ATTRACTIONS:
For local attractions, please visit the hotel website http://www.doubletreeplymouth.com/, or the Philadelphia Convention & Visitors Bureau website http://www.pcvb.org/.

SYMPOSIUM REGISTRATION (in USD):
(See later pages for details)

PROGRAM COMMITTEE:
CHAIR: *William Wei*, Temple University, (215) 204-8459 (v1000e@vm.temple.edu)

MEMBERS:
Danny Chaing, Janssen Research Foundation, (609) 730-3297 (dchaing@janus.jnj.com)
Ivan Chan, Merck Research Lab., (484) 344-3391 (ivan_chan@merck.com)
George Chao, Novirio Pharmaceuticals, Inc., (617)-250-3100 (chao.george@novirio.com)
Yusong Chen, AstraZeneca, (302) 886-5139 (yusong.chen@astrazeneca.com)
Alice Hsuan, Janssen Research Foundation, (609) 730-3001 (ahsuan@janus.jnj.com)
Francis Hsuan, Temple University, (215)-204-8105 (fhsuan@msn.com)
Lee Huang, Aventis Pharma., (908) 397-6413 (lee.huang@aventis.com)
Frank Shen, Bristol-Myers Squibb Co., (609) 818-6505 (frank.shen@bms.com)

SHORT COURSES PRESENTED AT ICSA 2002 Applied Statistics Symposium

June 6 (Thursday), 2002

Doubletree Guest Suites at Plymouth Meeting in Greater Philadelphia

- For extra information about the short course please contact: Ivan S.F. Chan, Merck Research Laboratories (484) 344-3391, Ivan_Chan@Merck.Com, or visit ICSA web site: *http://www.icsa.org*
- Registration to the Symposium program on 6/7-8 is highly encouraged but not required for short course participants.

Course 001: Design Considerations for Positive Late Phase Confirmatory Trials

(9:00 AM - 5:00 PM)

Instructors: ***Irving K. Hwang, Ph.D.***
University of Medicine & Dentistry of New Jersey, USA; Irving Consulting Group, USA

Course Outline:

Clinical trials in new drug development are classified into four Phases (Phase I-IV). Phases I-IIa trials are generally considered as early phase trials primarily for exploratory purposes. Phase III (sometimes Phase IIb) trials are late phase confirmatory trials. Phase IV trials are mostly post approval trials for long-term safety surveillance as well as for various marketing purposes. Trials of all phases are important for drug development, but the key for successful registration and approval for marketing lies in positive late phase (IIb-III) confirmatory trials. A positive confirmatory trial needs to demonstrate findings not only clinically relevant, but also statistically significant.

To improve the likelihood of success for new NCEs approval and launching, quality design and conduct of the late phase trials are imperative. This tutorial intents to provide insights in Good Clinical Practice (GCP) and Good Statistical Practice (GSP) in design considerations specifically for positive late phase confirmatory trials. Topics include hypothesis testing, statistical power, sample size estimation/adjustment, multicenter trials, superiority/non-inferiority trials, dose-response trials, interim analysis/group sequential designs/adaptive designs, and many other important trial design issues will be discussed. Several highly positive landmark megatrials will be given for illustration purpose.

About the Instructors:

Dr. Irving Hwang is currently President, Irving Consulting Group (ICG); Adjunct Professor, University of Medicine & Dentistry of New Jersey (UMDNJ). He was formerly PhRMA Deputy Topic Leader, ICH El0 Expert Working Group and Pharmaceutical Scientific Advisor, Bureau of Pharmaceutical Affairs, Republic of China. Dr. Hwang has gained over a quarter century of global drug development experience with major pharmaceutical companies in design and analysis of clinical trials for development of new drugs and

vaccines. Previously, Dr. Hwang was Sr. Vice President, Harvard Clinical Research Institute; Vice President & Head, Global Biometrics, Hoechst Marion Roussel, Inc. (HMR); Sr. Director, Clinical Research Operations, Hoechst Roussel Pharmaceuticals, Inc. (HRPI); and Sr. Director, Clinical Biostatistics & Research Data Systems, Merck. He was Member, PhRMA BSS Steering Committee and Co-Chair, PMA/FDA Workshop on Clinical Trials Monitoring and Interim Analysis.

Dr. Hwang received his Ph.D. in Statistics from the Wharton School, University of Pennsylvania. His research interests include PK/PD modeling, regression analysis, robust methods, survival analysis, longitudinal analysis, interim analysis/group sequential methods, and confirmatory clinical trial methodology including design and analysis of landmark megatrials and non-inferiority/equivalence trials.

Course 002: Active Control Non-Inferiority Studies and Adaptive Analysis Methods in Clinical Trials

(9:00 AM - 5:00 PM)

Instructor: *Sue-Jane Wang*[1], Ph.D. and ***H.M. James Hung*[2], Ph.D.,**
[1]Division of Biometrics II, [2]Division of Biometrics I, Office of Biostatistics, Center of Drug Evaluation and Research, Food and Drug Administration

Course Outline: The first topic of this course is non-inferiority testing. In an active controlled non-inferiority trial without a placebo arm, 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 course 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 second topic is adaptive design/analysis. The success of a clinical trial relies on many design specifications that often need to be accurately projected. A notable example is that the effect size of an experimental treatment needs to be accurately predicted in order that the sample size can be planned to ensure a sufficient power for detecting the treatment effect. Another example is that projections are needed to design a small number of primary endpoints and distribute the total alpha for testing them. As such projections depart greatly from the true state of nature, the costly clinical trial has a high probability of failure. In the past few years there has been a growing demand of enhancing the flexibility of clinical trial design. In particular, often an attempt is made to modify some design elements using the information collected from the accumulating data at some interim time as the clinical trial proceeds. The conventional statistical test may be invalid as a result of such modification. A possible viable solution to this problem is adaptation of the conventional test procedure as proposed by Cui, Hung and Wang (Biometrics, 1999) for sample size re-estimation. We shall illustrate how the adaptive test is implemented so that sample size can be adjusted without compromising the validity of statistical inference in group sequential and non-group sequential trials. Such adaptation will be extended to handling multiple endpoints and to implementing a closed test procedure for testing both superiority and non-inferiority in active controlled trial. Issues on the operational aspect of such adaptation will be discussed.

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. H.M. James Hung is currently an acting team leader and an expert mathematical statistician in CDER, FDA. He received a FDA level CDER Scientific Achievement Award. Dr. Hung received his Ph.D. in Statistics from Iowa State University in 1983.

Drs. Hung and Wang and two other colleagues in a RSR team received a FDA Award of Merit for the research achievement on adaptive design/analysis. Dr. Wang is Principle Investigator for the RSR project on active control non-inferiority trials and Dr. Hung is Principle Investigator for adaptive design/analysis.

Course 003: Statistical Approaches in Pharmacogenomics

(9:00 AM - 5:00 PM)

Instructor: *Kim E. Zerba, Ph.D., Shu-Pang Huang, Ph.D., and C. Frank Shen, Ph.D.*, Pharmaceutical Research Institute, Bristol-Myers Squibb

Course Outline: Pharmacogenomics is a hybrid discipline that seeks to discover genes whose physiologic influences may lead to the development of better drugs. A hierarchical, complex adaptive systems approach to questions consistent with the role of genetic variation in biological variation provides the foundation for successful pharmacogenomics studies. Such an approach considers the influence of genetic variation on disease endpoints, surrogate intermediate physiological and biochemical biomarkers, treatment efficacy and adverse events in proof-of-principle clinical trials. The design of and interpretation of inferences from these trials simultaneously invokes principles in genetics, genetic epidemiology, molecular biology, biochemistry, physiology, evolutionary biology and biostatistics. This short-course will be presented in five parts and will focus on some of the key genetic,

statistical and clinical issues relevant to implementing pharmacogenomics studies. The course will provide the basic genetics perspective and background knowledge necessary for understanding, and approaches to resolving, these key issues. The course will end with an examination of additional biological, statistical, and ethical challenges that remain. The following outline represents the topics that will be covered.

- **The Role of Genetic Variation in Pharmacogenomics and Drug Development**
- **Genetics Perspective**
  - General/Historical Genetics Paradigm
  - Molecular Genetics
  - Population Genetics
  - Phylogenetics
- **Statistics Perspective**
  - Case-Control Sampling Design Focus for Pharmaceutical Industry
  - Genetic Variation
    - Population Genetic Structure
    - Linkage Disequilibrium
    - Traditional Statistical Approaches
- **Clinical, Regulatory and Marketing Perspective**
  - Data Privacy
    - Anonymization/De-Identification
      - Genetic Sample Banking
- **Key Statistical Issues and Approaches for Pharmacogenomics Studies**
  - Admixture of Gene Pools from Different Ethnic Groups
    - Genomic Control
  - Linkage Disequilibrium
    - Haplotype Reconstruction
    - Cladistics
    - Combinatorial Partitioning
  - Multiplicity

About the Instructor: Dr. Kim Zerba is currently Associate Director of Statistical Genetics and Biomarkers for Clinical Discovery Biostatistics and Data Management at Bristol-Myers Squibb Co. He received his Ph.D. in Zoology, specializing in evolutionary biology, ecology and biostatistics, from Arizona State University in 1989. His postdoctoral research focused on statistics, genetics and the role of genetic variation in complex human diseases, particularly cardiovascular disease, in the Department of Human Genetics at the University of Michigan until 1993. He then continued with the University of Michigan as a research scientist conducting research on the genetics of complex diseases until 1999 when he joined the Bristol-Myers Squibb Pharmaceutical Research Institute.

Dr. Shu-Pang Huang is currently a biostatistician in the Statistical Genetics and Biomarkers group in Clinical Discovery Biostatistics and Data Management at Bristol-Myers Squibb Co. Shu-Pang received his Ph.D. in Statistical Genetics under the supervision of Bruce Weir from North Carolina State University in 2001. His dissertation research interest focused on estimating the number of rare alleles and their frequencies. At Bristol-Myers Squibb, he has been directly involved in pharmacogenomics studies and applying the statistical genetics approach to case-control samples from large clinical trials.

Dr. Frank Shen is currently Head of Clinical Discovery Biostatistics and Data Management at Bristol-Myers Squibb Co. and has been leading a group of statisticians to fulfill a mission of integrating all statistical collaborations in drug discovery, statistical genetics, pharmaceutical development, clinical Pharmacology, and human pharmacokinetics into a close partnership between scientists and statisticians. He is also instrumental in participating several company-wide initiatives including productivity improvement, electronic dossier submission, and pharmacogenomics. Frank joined the industry in 1989 and worked at Wyeth-Ayerst Research as a research statistician until 1993. He joined the Biometrics Research group at Merck & Co. in 1993 and was manager of the Rahway group prior to joining Bristol-Myers Squibb in 1996. Frank received his master degree in Chemical Engineering from Lamar University, Texas, and Ph.D. in Statistics from Temple University.

Course 004: Advanced Log-linear Modeling for Categorical Data With GENMOD

(9:00 AM - 5:00 PM)

Instructors: ***Daniel Zelterman, Ph.D.***
Division of Biostatistics, Yale University

Course Outline:

The SAS GENMOD procedure is a flexible implementation of the Generalized Linear Model. In this tutorial, we will see how GENMOD can be used to fit a wide variety of problems in categorical data. We begin with a brief discussion of discrete distributions and log-linear models in elementary settings such as models for 2x2 tables. The first half of the session emphasizes log-linear models for non-rectangular tables and tables with ordered categories. Examples include a circular table of frequencies indicating locations of brain lesions in stroke patients. Another class of problems we discuss are mark-recapture settings in which we need to estimate the mean of an unobservable frequency in a multi-dimensional table. The second half of the tutorial describes extensions to GENMOD that allow us to develop models for new sampling distributions. One sampling distribution is a truncated Poisson model in which the zero frequencies are not observed. The GENMOD procedure is provided to perform truncated Poisson regression. Another sampling distribution is the hypergeometric model for 2x2 tables. The log odds-ratio parameter for these tables can be modeled using a hypergeometric regression that can be performed in GENMOD. All of the programs and examples are taken from the speaker's forthcoming book to be published in the SAS Books by Users series. An additional text book for the course is "Models for Discrete Data" published by Oxford in 1999.

About the Instructors:
Dr. Daniel Zelterman completed his Ph.D. in Statistics at Yale in 1983. He is Professor in the Division of Biostatistics at Yale. He has also been on the faculty at the State University of New York at Albany, and at the University of Minnesota. He is the Director of the Biostatistics Core at the Yale Cancer Center and serves as Associate Editor of four journals including Biometrics. He was recognized as Fellow of the American Statistical Association in 1998. His scholarly output includes over 100 articles and two books including "Models for Discrete Data" published by Oxford in 1999.

INTERNATIONAL CHINESE STATISTICAL ASSOCIATION
YEAR 2002 APPLIED STATISTICS SYMPOSIUM

Registration Form

JUNE 6-8, 2002

Doubletree Guest Suites at Plymouth Meeting in Greater Philadelphia

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

Affiliation: _____

Affiliation category: Industry _____ Government _____ Academic _____

Mailing Address: _____

_____

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

Symposium Registration (in USD. Please check the appropriate box):

Membership Type	By April 30, 2002	Check (√)	After April 30, 2002	Check (√)
Regular Member	$100		$120	
Regular Nonmember	$140		$160	
Student Member	$30		$40	
Student Nonmember	$50		$60	

Note: The symposium registration fee includes the program package, breakfasts, and lunches on June 7-8

ICSA Membership for 2002:

Membership Type	Check (√)
Annual ICSA regular membership: $40	
Annual ICSA student membership: $20	
Lifetime ICSA permanent membership fee: $400	

Note: New members need to fill out an ICSA Membership Application Form (which can be downloaded from http://www.icsa.org)

Short Courses (Thursday, June 6, 2002):

Topic	Time	Instructor
Design Considerations for Positive Late Phase Confirmatory Trials	9:00 AM – 5:00 PM	Dr. Irving Hwang University of Medicine & Dentistry of NJ and Irving Consulting Group
Active Control Non-Inferiority Studies and Adaptive Analysis Methods in Clinical Trials	9:00 AM – 5:00 PM	Drs. Sue-Jane Wang & James Hung CDER, FDA
Statistical Approaches in Pharmarcogenomics	9:00 AM – 5:00 PM	Drs. Kim Zerba, Shu-Pang Huang & Frank Shen, Bristol-Myers Squibb
Advanced Log-linear Models for Categorical Data With GENMOD	9:00 AM – 5:00 PM	Dr. Daniel Zelterman Yale University

If you have chosen to attend one of the short courses, please check the fee you need to pay:

Membership Type	By April 30, 2002	Check (√)	After April 30, 2002	Check (√)
Regular Member	$250		$300	
Student Member	$125		$150	

Note: The short course registration fee includes breakfast, lunch, and coffee breaks.
A short course may be canceled due to lack of participation. This fee will be fully refunded in such a case.

Banquet: ($35 + _____ × $25/each additional family member) $_____

Donation: _____ $_____

Total Amount Included: $_____

PLEAE SEND REGISTRAION FORM WITH CHECK (Payable to ICSA) TO:
Yusong Chen
Treasurer, ICSA 2002 Applied Statistics Symposium
AstraZeneca LP
1800 Concord Pike
PO Box 15437
Wilmington, DE 19850-5437
Phone: (302) 886-5139
Fax: (302) 886-1099
E-Mail: yusong.chen@astrazeneca.com

Cancellation Policy: Unless approved by the Committee, all symposium participants must register except for keynote, dinner, and plenary speakers. Full refund for cancellation made on or before April 30; 80% refund after April 30 but on or before May 15.

HOTEL INFORMATION

Are you going to stay in DoubleTree Guest Suite? Yes _____ No _____

Participants should book their own hotel reservations. Special group rate $119 per suite for one or two people at the Doubletree Guest Suites with reservation made no later than May 10, 2002. The number of rooms is limited on a first-come-first-served basis. Be sure to make your reservation early, as the hotel is extremely busy. It allows cancellation any time before 4 p.m. on the date of your arrival. To make a reservation, please call 610-834-8300 or 800-222-8733. The mailing address is as follows:

DoubleTree Guest Suites Plymouth Meeting
640 West Germantown Pike
Plymouth Meeting, PA 19462

Call for Nomination for President and Board Directors

The ICSA Nominating and Election Committee invites you to nominate candidates for the 2003 President-elect and five members of Board Directors. Your input is important as the success of ICSA depends on strong leadership which can be best produced from a broad pool of candidates of these two offices. We are looking forward to receiving nominees from you. Please check with nominees about their qualifications before you submit their names to Naitee Ting, the Committee Chair. Nomination deadline is February 28, 2002. 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.

Nominating and Election Committee:
Jeff C. F. Wu, Jen-Pei Liu, Heping Zhang and Naitee Ting (Chair)

Please submit your nomination to:

Naitee Ting
MS 6025-C2258
Pfizer Global Research & Development
50 Pequot Ave.
New London, CT 06320
phone (860) 732-4871, fax (860) 732-7806
Emai: naitee_ting@groton.pfizer.com

Call for Nomination of New ICSA Bulletin Editor

Editor-in-Chief
(2003-2005)

The publication Committee invites ICSA members to nominate a new Bulletin Editor. The new BE will work with Dr. Sue-Jane Wang for the July 2002 issue and will assume the responsibility for editing the ICSA Bulletins and coordinating with various ICSA officers to bring necessary reports to publication starting January 2003.

The qualifications for a candidate are

- Current ICSA member
- Commitment to work as a Bulletin Editor
- Unbiasedly selecting and publishing the submitted member news

Please send your nomination letter to the Publication Committee by March 31, 2002. Before you nominate a member, make sure that the person accepts the nomination. In your letter, please provide one or two paragraphs describing the nominee s qualification as a candidate for the BE. The nomination maybe submitted by any one of the following delivery modes:

By E-mail, jchen@nctr.fda.gov
By FAX, 1-870-543-7662 (To James J. Chen, Ph.D.)

Publication Committee
James J. Chen
I-Shou Chang
Jun Shao
Sue-Jane Wang
Ker-Chau Li
Yi Tsong

Nominations for 2002 ICSA Awards

The ICSA Awards Committee invites you to nominate candidates for the 2002 ICSA Awards. The committee is looking for "a few good members" who have made significant contributions to the Association. Examples of candidates are members who have:

1. Successfully organized and conducted a professional meeting
2. Effectively led the ICSA activities
3. Held a strong and effective editorship (including associate editorship) for Statistics Sinica or ICSA Bulletin
4. Substantially increased the ICSA membership
5. Initiated and implemented a policy which resulted in a substantial visibility for the ICSA
6. Established strong ties with other professional societies and co-sponsored statistics activities with them.

In your nomination please include the nominee's name, title, address, a description of his/her significant contributions and the names of other ICSA members or statisticians who are familiar with the nominee's contributions. Once the awards are decided, they will be presented in August 2002 in New York.

The committee will select the most deserving candidates, who have made multiple significant contributions resulting in a strong positive impact to the Association, to receive awards. Preference will be given to those who have not previously received such an award.

Your nominations must be received no later than February 28, 2002. Please send your Nominations to:

Lynn Kuo
Chair, ICSA Awards Committee
c/o Department of Statistics
University of Connecticut
Storrs, CT 06269-4120

REGIONAL ACTIVITY

Some Upcoming Statistical Meetings in Taiwan

by C. Andy Tsao

The 2002 Taipei International Statistical Symposium and Bernoulli Society EAPR Conference
7--10 July 2002, Academia Sinica, Taipei, Taiwan.
URL: http://www.stat.sinica.edu.tw/2002symp/

The 2002 Southern Taiwan Statistical Conference
27—28 June 2002, National Sun Yat-Sen University, Kaosiung, Taiwan.
URL: http://www.math.nsysu.edu.tw/conference/ssc2002/

New Trend, New Organizations and New Name
In the knowledge-based society, organizations play more important roles than they used to be. New jobs await to be done and call for new organizations/teams. Here are some new organizations emerging in Taiwan.

Chunghwa Data Mining Society (CDMS)
Data mining is an important component of Information Technology. This society is founded in December 2001. Its mission is to develop data mining technology, to integrate the data analysis methodology, database technology and applications, to provide decision-making information and to facilitate knowledge-based economic.

Division of Biostatistics and Bioinformatics at National Health Research Institutes
Founded in 1997, the institute, Division of Biostatistics then, was the first academic biostatistics institution in Taiwan. Due to the growing importance of bioinformatics in the genomic Research at NHRI, starting from December of 2000, the Division is now the Division of Biostatistics and Bioinformatics.
URL: http://www.nhri.org.tw/nhri_org/bs/biostat/e_intro.htm

Sinoacademic Survey Network (SASN)
Technological innovation is essential to growth of output and productivity. In Taiwan, the first technological innovation survey is now ongoing. The project is jointly funded by National Science Council (Taiwan) and Ministry of Economic Affairs. The survey team, SASN, consists of statisticians, subject field scientists from universities and research institutes. Through the cooperation, they seek to standardize the survey conducting, harmonizing education and training for survey practitioners.
URL: http://140.136.12.12/ (SASN)

Department of Statistics and Information Science at Fu-Jen University
In recognizing statistical science as an essential part of information technology, this department now has a new name, new identity and new direction.
URL: http://www.stat.fju.edu.tw

Report of Hong-Kong Regional Activity/news in 2001

By Hailiang Yang

Here I report some of the activities and news of 2001 from the Hong Kong statistical community. I apologise for any missing information.

Professor Howell Tong won the National Natural Science Award.

Professor Howell Tong, Chair of Statistics and Actuarial Science Department of the University of Hong Kong, has been awarded the National Natural Science Prize (China) for his contribution in non-linear time series analysis. Professor Tong was the only winner from Hong Kong across all disciplines and the only mathematician winning the award in 2000. Earlier, in February 2001, the prize was presented at a ceremony in Beijing, attended by major state leaders.

The 5th ICSA International Conference

The 5th ICSA International Conference was held at the University of Hong Kong in August 2001. For details, see the report in this issue.

Launching of the HKSS Professional Examination

The Hong Kong Statistical Society will formally launch its professional examination as from May 2002. This will replace the examination held by the Royal Statistical Society in Hong Kong. This will be the first statistical professional examination offered by a statistical society in the Asian region. To promote the examination, a press conference was held on October 3, 2001. A press release was issued to the media on the same day.

Statistical Seminar

Invited by the Hong Kong Statistical Society, Professor Stephen Fienberg, Professor of Statistics and Center for Automated Learning and Discovery, Carnegie Mellon University, presented a talk entitled "The Interplay between Research Innovation and Federal Statistical Practice" on 21 November 2001.

Promotion

Professor W. K. Li, Department of Statistics and Actuarial Science, the University of Hong Kong, has been appointed as Chair Professor of Statistics.

Drs. Minggo Gu and Wai Yin Poon, Department of Statistics, The Chinese University of Hong Kong, have been promoted to Reader since October 2001.

STATISTICS' DELIGHT
統計趣聞

統計謎語

- 如果統計是從有限的資料、線索尋找答案的一門學問，則謎語也可以看做是統計問題。

根據以下的提示，射一統計相關語。

1. 三日入廚下，洗手作羹湯，
   未諳姑食性，先遣小姑嘗。
   (語出唐詩)
2. 但見新人笑，那聞舊人哭。
3. 前不見古人，後不見來者。
4. 苦恨年年壓金線，
   爲他人做嫁衣裳。
5. 考前猜題，考場作弊。
6. 離離原上草，一歲一枯榮，
   野火燒不盡，春風吹又生。
7. 循序漸進。
8. 滄海一粟。
9. 一人之下，萬人之上。
   (猜 ICSA Past President)
10. 戲法人人會變，巧妙各有不同。

答案

1. Surrogate Endpoint
2. New Better Than Old
3. Dirac Point Mass Function
4. Clinical Statisticians
5. Forecasting & Interim Look
6. Renewal Process
7. Multi-Stage Design
8. Statistically Insignificant
9. 傅權，永遠是ㄈㄨˊ總統
10. How to lie with Statistics

Hypothesis ???

☺☺☺

"I cannot give any scientist of any age better advice than this: the intensity of a conviction that a hypothesis is true has no bearing over whether it is true or not." - Peter Medawar (1915 - 1987), zoologist, winner of the Nobel Prize in medicine.

☺☺☺

Do you know ???

Statistics - A subject which most statisticians find difficult but in which nearly all physicians are experts.

Medical Statistician - One who won't accept that Columbus discovered America because he said he was looking for India in the trial plan.

Equivalence Trials - Proving that apples are pears by comparing their weight.

Sequential Analysis - A means of stopping a trial before it becomes useful.

Trend towards Significance - An ever-present help in times of trouble.

"Truth does not change because it is, or is not, believed by a majority of the people." -Giordano Bruno
"Poker is a game of chance, but not the way I play it." - W.C. Fields

The devil was instructing some new recruits on some of his maxims. 'There are three ways to misrepresent the truth', he said. 'A lie, a damn lie, and statistics!'

CANDA: A submission, which is printed out by the FDA rather than by the sponsor.

Bayesian: One who, vaguely expecting a horse and catching a glimpse of a donkey, strongly concludes he has seen a mule.

There are three kinds of lies: lies, damned lies, and statistics.
- Benjamin Disraeli (1804 - 1881)

Cigarette smoking is a major cause of statistics. -Unknown

Mathematics is the art of giving the same name to different things.
--Jules Henri Poincare

"Obvious" is the most dangerous word in mathematics.-- Eric T. Bell

The great tragedy of science -- the slaying of a beautiful hypothesis by an ugly fact.
--Thomas Huxley

The scientific theory I like best is that the rings of Saturn are composed entirely of lost airline luggage. --Mark Russell

What we see depends mainly on what we look for. --Sir John Lubbock

New ideas pass through three periods:
*It can't be done.

*It probably can be done, but it's not worth doing.

*I knew it was a good idea all along ! -- Arthur C. Clarke

From the Desk of the Editorial Working Committee

Welcome New Members

(Joined between October 2000 and Dec. 2001)

We would like to welcome new members who joined between November 2000 and December 2001.

Michael P. Cohen
Chunming Zhang
William W. Chen
Dale Song
Hongquan Xu
Runze Li
Bann-Mo Day
Guan-Hua Huang
Juan Zhang
Liang Zeng
Hongyu Wu
Huaixiang Li
Yuanjie Michael Zhang
Kamchuen Su
Xiangrong Yin
Qiang Zhao
Yichuan Zhao
Chen Yang
Paul Pathouz
Gang Li
Peter Michael Thompson
Joel A Dubin
Katherine Chi-Burris
Klaus Poetzel Berger
Quanxi Shao
Shao-Wei Cheng
Erica Hua Gao
Chia-Wen Ko
Chiu-Hsing Weng
Sujuan Gao
Karl Heiner
Ping-Hung Hsieh
Marven Zelan
Bi-Min Hsu
Jie Chen
Fritz Scheuren
Zhaohui Qin
Yu Shyr
Ying Chen
Yazhen Wang
Yuguo Chen
Hee-Jing Zhong
Yuqing Dai
Yunfan Deng
Zhen Luo
Gibb Bassett
Jing Wang
Zheng Yuan Zhu
Na Li
Pei-Whe Hsu
Ming-Yi Hu
Lei Shen
Heng Li
Shuo-Jye Wu
Yi-Ting Huang
Hock Peng Chan
Mingxiu Hu
Meiyu Shen
Samuel Kou
Dean Slonowsky
Zhiqiang Tan
John Tuhao Chen
Lu-Sa Su
Erin M. Hodgess
Phuoc Le
Geng Sheng Qin
Weng-Yao Ku
Zhe Li
Xu Yan
Xiao Tong Shen
Xiang Qui Qu
Yufei Huang
Ye Ding
Richard Zitikis

Congratulations to the Following ICSA Members Who Were Elected as ASA, IMS, and COPSS Fellows or Receiving Awards in 2001

Prof. George Tiao received the ASA 2001 Wilks Memory Award and Shiskin's Award

Prof. Xiao-Li Meng received the 2001 COPSS Award

Profs. Kai-Tai Fang, Jason Hsu, Kung-Jong Lui, Wei Yann Tsai, and Naisyin Wang were elected as new ASA Fellows

Prof. Qi-Man Shao was elected as a new IMS Fellow

北京大學光華管理學院

Guanghua School of Management
Peking University

Faculty position openings
in
Department of Statistics and Econometrics

Guanghua School of Management at Peking University, a leading business school in China, is establishing a new **Department of Statistics and Econometrics**. We are conducting a worldwide recruiting for full time faculty at all ranks (short term visiting positions are also available.) Applicants should have a Ph.D. degree in statistics or econometrics and an established academic record or strong research potential. Those with experience and tenure in prestigious universities or research institutes in Europe and North America are especially welcome.

The School offers undergraduate, MBA, EMBA, M.S. and Ph.D. programs, with an enrollment of 2,000 students, of whom 1,200 are in the MBA program. The School is partially endowed with US$10 million by the Guanghua Education Foundation (HongKong). Please visit *www.gsm.pku.edu.cn* for more information.

The starting salary for assistant, associate and full professors are US$40,000, $50,000 and $60,000, respectively (all payable in Chinese currency). In addition to the standard salary and fringe benefits, successful candidates for full-time faculty positions will also receive the following special benefits.

1. Housing benefits. The university provides a two- or three-bedroom apartment; those who wish to buy their own apartment will receive a lump-sum subsidy of 200,000 RMB.
2. Settle-down allowance of 150,000RMB.
3. A start-up research grant of 50,000RMB per year for the first two years.
4. Passage allowance. The School will reimburse one-way economy class airfares for the appointee and his/her immediate family members to fly to Beijing for duty.
5. Flexible sabbatical leave. Sabbatical leaves can be arranged for faculty with well-established records and previous appointments in overseas academic institutions.
6. Publication reward: the current reward for each English publication varies from 10,000RMB to 40,000RMB (depending on journal ranks).

Please contact Professor Rong Chen at (US)312-996-2323 for more information. Applicants please send a CV, three reference letters and a statement on teaching and research to Professor Rong Chen, Department of Information and Decision Sciences (MC 294), College of Business Administration, University of Illinois at Chicago, 601 South Morgan Street, Chicago, IL 60607, U. S. A.

International Chinese Statistical Association
Profit & Loss
January through December 2001

Ordinary Income/Expense	
Income	
Banquet at ASA Meeting	2,720.00
Membership Dues	15,650.00
Miscellaneous Income (short course)	173.00
Total Income	18,543.00
Expense	
Bank Service Charges	0.40
Board Meeting	808.00
Computer Hardware/Software	
Laptop PC	1,798.78
Other Hardware/Software	812.00
Total Computer Hardware/Software	2,610.78
Contributions	
ASA	500.00
Total Contributions	500.00
ICSA at ASA meeting	
Banquet	2,491.00
Total ICSA at ASA meeting	2,491.00
Licenses and Permits	95.00
Postage and Delivery	
Announcement	338.99
Ballot	1,120.99
Book/Journal Donation	2,307.38
Bulletin	3,460.04
Other	692.32
Total Postage and Delivery	7,919.72
Printing and Reproduction	
Jan. Bulletin	3,921.00
July Bulletin	3,957.00
Total Printing and Reproduction	7,878.00
Professional Fees	
Tax filing (1998, 1999, 2000)	750.00
Total Professional Fees	750.00
Supplies	
Office	1,158.06
Other	941.53
Total Supplies	2,099.59

Travel	
Travel	331.84
Total Travel	331.84
Total Expense	25,484.33
Net Ordinary Income	-6,941.33
Other Income/Expense	
Other Income	
Interest Income	1,476.88
Total Other Income	1,476.88
Other Expense	
1998 & 1999 tax	2,715.00
Total Other Expense	2,715.00
Net Other Income	-1,238.12
Net Income	-8,179.45

International Chinese Statistical Association

Balance Sheet

As of December 31, 2001

ASSETS	
Current Assets	
Checking/Savings	
checking	7,581.02
Savings-CD	30,694.49
Savings-Money Market	35,479.20
Total Checking/Savings	73,754.71
Total Current Assets	73,754.71
TOTAL ASSETS	73,754.71
LIABILITIES & EQUITY	
Equity	
Opening Bal Equity	81,934.16
Net Income	-8,179.45
Total Equity	73,754.71
TOTAL LIABILITIES & EQUITY	73,754.71

INTERNATIONAL CHINESE STATISTICAL ASSOCIATION

Membership Application / Renewal Form (2002)
Date _____ New Member _____ Renewal _____

NAME (Last) (First)

English:

Chinese:

ADDRESS

Office:	Home:
City: State:	City: State:
Zip:	Zip:
Country:	Country:
Tel (O):	Tel (H):
Fax (O):	Fax (H):

E-Mail Address:

Highest Degree: Year Graduated:

University: Occupation/Title:

MEMBERSHIP FEES

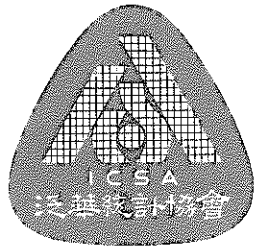
Regular	US$40	_____
Student	US$20	_____
Permanent	US$400	_____
Spouse (50%)		_____ (Spouse Name _____)
Biometrics (Free)		_____
Donations		_____
Total Amount		_____

STATISTICAL AREA OF INTEREST (circle as many as you like):

A. Agriculture
B. Business/Econometrics
C. Computing/Graphics
D. Education
E. Engineering
F. Health Sciences
G. Probability
H. Social Sciences
I. Theory And Methods
N. Biostatistics

Please make checks payable to I.C.S.A.

Mail this form and a check to: I.C.S.A.
c/o Yi Tsong, Ph.D.
13215 lazy Glen Lane
Herndon, VA 22071
U.S.A.
(tsong@fda.cder.gov)



INTERNATIONAL CHINESE STATISTICAL ASSOCIATION

Information Sheet (2002)
Date _____

RECENT NEWS: (publications, research or teaching activities, job transfer, awards or honors received, etc.)

I. C. S. A.
13215 Lazy Glen Lane
Herndon, VA 22071
U.S.A.

JIA-YEONG TSAY (P)
Organon Inc., 375 Mt. Pleasant Avenue
WEST ORANGE NJ 07052
United States