

# BIOSTAT CONNECTIONS

Editorial Staff (2001): Victor DeGruttola, Brent Coull, Marcia Testa, Florin Vaida,  
Bhaswati Ganguli, Nora Horick, Evelyn Ophir  
Editorial Staff (2002): Rebecca Betensky, Karen Kuntz, Mei-Chiung Shih, Matt Wand,  
Jelena Follweiler

July 2001/2002

No. 11

## An Interview with Professor Wing Wong by Florin Vaida

Professor Wing Wong has joined the Department of Biostatistics in the Fall of 2000. After having made an important impact in various areas of statistics – recognized, most notably, by the 1993 COPSS Medal, the most prestigious award in Statistics, offered yearly to a statistician under the age of 41, for the past several years he has been working full-time in the newly emerging area of Computational Biology. He is now spearheading our department's strategic effort to be a leader in this very dynamic field. Wing is "a statistician's statistician". Modest to the point of self-effacement, looking ten years less his age, you can easily take him for a post-doc or a graduate student on the busy hallways of the department. However, it's enough to talk to him or read his papers to be impressed by his depth of thought, clarity, and commitment to science. Wing is also one of my personal heroes. Back in 1994 at the University of Chicago I took a class on Markov Chain Monte Carlo from him, and I decided then that this is the area I wanted to work in. I've acknowledged him on my dissertation, on the same page with my thesis advisor and with Michael Jordan. I welcomed therefore the opportunity of the following interview for the "Newsletter".

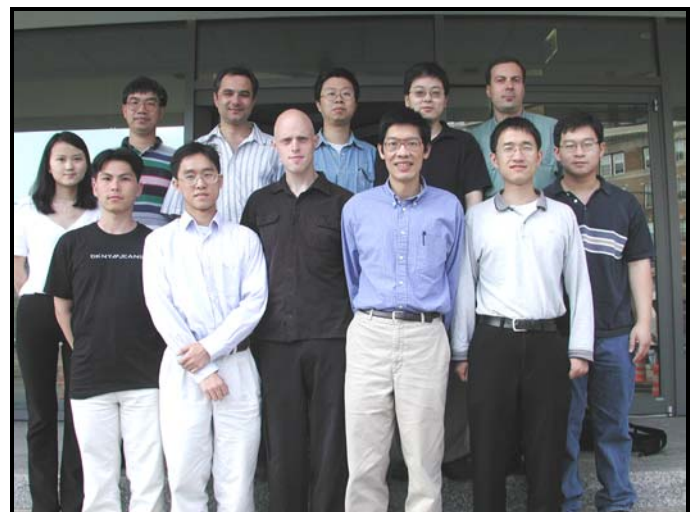
FV: Wing, you have a Ph.D. in Statistics and a M.S. in Computer Science. What are you? A statistician, a computer scientist, or a biologist? Is a Biostatistics department a natural place for your research?

WW: I tend not to think in these terms, I like to think of myself as a scientist, and apply whatever tools I have available to the research questions of interest to me. For the moment it is mainly biology questions, and how to analyze the large amount of data that is pouring out of these high throughput biological experiments, but it is conceivable that in the future my interest will take another turn – it happened several times in my career. I think that Statistics offers a wonderful preparation for this type of activity, we are very comfortable with data, with interacting with scientists and with thinking about various kinds of problems, and it is relatively easy for us to take advantage of new opportunities. Computational Biology is an area that any department can claim as its own, and in the long run the discipline that is most aggressive and

most receptive will claim the biggest prize. This happened again and again: Statistics has been competing with Computer Science and Applied Math for many of the mathematical and computational problems that arose in all sorts of scientific areas. It is up to us to pursue these problems, and if enough of us are doing this, then they will become a part of Statistics.

FV: More specifically, what problems are you working on these days?

WW: We spend a lot of time thinking how to make use of the new microarray technologies and anticipated high-throughput proteomics technology in various applications to biometrical research. For example, very large projects are going on now in the Harvard medical area on the large scale applications of gene-array technology to look at clinical samples related to cancer and other diseases. These kinds of problems are very statistical, we work with groups in Dana Farber and in the



From left to right: (back row) Xianghong Zhou, Horng-Shing Lu, Ovidiu Lipan, Xuegong Zhang, Yan Cui, Igor Leykin; (front row) Han-Ming Wu, Chien-Cheng Tseng, Byron Ellis, Dr. Wing Wong, Cheng Li, Ming-Chih Kao.

Wong Interview

(cont'd next page)

Wong Interview

(cont'd from prev page)

medical school that generates large amounts of data, hundreds and thousands of arrays at a time and try to correlate the phenotype information, in many cases including treatment history, with the information in the gene expression array. There are many challenges. Another area we are pursuing very seriously regards functional genomics approaches to very central cell biology questions, such as processes and pathways relevant to cell cycle control, cell proliferation, cell death, disease development, and so on. Developmental biology is a really central area of modern biology, with dramatic advances in the past ten years; we are working with some of the developmental biology groups in the campus on stem cells for example, trying to understand what makes a cell a stem cell, what is the sequence of events that trigger a particular developmental pathway. Admittedly, we are newcomers to this area and we have lots of biology to learn. We are lucky to have scientists close-by who generate the data or ask us to analyze them, and we take this opportunity to learn some biology and try to contribute the best we can. It is not clear what approaches to take. There are groups that adopt very holistic approaches, "we'll use a cyclic digraph or a Bayesian network". We tend to not start with a preconceived idea on how to approach this problem. Rather, at this point we are just taking our time observing how the biologists think about this and we do what we can to help them make progress, and I think hopefully in next one-two years we will get enough understanding to start formulating a more systematic analytical approach; of course, we are keeping track of what other groups are doing. Another important research activity is the development of software tools to facilitate better data analysis in the field of functional genomics.

FV: Is statistical inference suited for these sorts of problems? It seems that the nature of the analysis is more exploratory.

WW: Statistics is definitely not the answer to all the questions in the area; I'd say it is an essential component of our approach, we have to somehow consider the statistical inference aspect, which means that when you get a result you have to think, is this real or just an artifact? And this is really statistical inference in the most general sense, and being trained as a statistician is an advantage.

FV: How are you fitting in the department so far? It sounds like you are already developing collaborations around here.

WW: The department so far has been extremely supportive, it's a very happy department, forward looking, and I think I'm on the same wavelength with most of my colleagues. Although we are into this at various degrees, I think almost everyone in the department is actively thinking about how traditional activities in biostatistics could benefit from the molecular biotechnology revolution that is unfolding. Of course, how this will affect clinical trials for example is a different set of problems from how it may affect basic biological research, but the overall direction is to take advantage of the availability of molecular level fundamental information about living organisms, including us humans. I think we are on the same wavelength!

FV: Why did you move from LA, was it the weather?

WW: No, I prefer LA weather [laugh again], but I see better scientific opportunities here which match very well with what my group wants to achieve.

FV: Can you tell us more about your group? You didn't come alone, you brought along a number of post-docs and graduate students.

WW: Last year five of us moved from LA, but now we are at about 10 people, including postdocs, graduate students and visitors. That's about the right size.

FV: I'd like to talk a bit about your career so far. You've worked on everything: asymptotics, nonparametrics, smoothing, survival analysis, Bayesian inference, Markov chain Monte Carlo, both theory and applications. What is the common theme here?

WW: No, by no means, I didn't work on everything! This is a hard question, I don't know if there is a common theme. If I happen to be thinking about a problem or I meet by chance collaborators with whom I can interact well in a problem that is of interest to me, I will think about that; I don't think it is a very well planned and designed path...

The common theme may be, my philosophy is not to think of boundaries of disciplines, not to be afraid to take on problems that apparently are very far from what you are doing at the moment, as long as you see a scientific value and have a genuine interest in it.

FV: How can one maintain depth and originality in several areas? Is it better to be broad, or to be focused on a smaller area?

WW: I don't know, I think there are arguments for both approaches. No question, if you spread yourself on multiple areas, your ability to contribute very deeply into any one of them is diminished, but I also believe that if you choose your problems correctly, at the right moment, you can make a very significant impact without having to solve the "deepest" problem in that area. Sometime the deepest problem in a mathematical, methodological sense may not be the most relevant! To have an overall, broad view of the scientific landscape is rather important in modern science, because things move so rapidly.

FV: Is there any particular contribution of yours that you are most proud of? The 1987 JASA paper with Martin Tanner was particularly recognized as bringing MCMC into Statistics several years before others...

WW: Yes, that was a good paper and I've very much enjoyed working with Martin on that. Another paper that I spent quite a bit of time on, in fact a pair of papers, were in the "Annals" in '94 and '95 with Xiaotong Shen, clarifying the rate of convergence of various nonparametric procedures, especially sieve-based methods. The '95 paper gave what I hope to qualify as a pretty clear picture of the behavior of the likelihood ratio surface in infinite dimensional situations. Technically, that was the most difficult work that I've done. I'm hoping that that approach will find more applications as time goes on in areas such as machine learning; that remains to be seen.

Wong Interview

(cont'd next page)

Wong Interview

(cont'd from prev page)

FV: I see you here all the time on Saturdays and Sundays. Is the weekend rest overrated?

WW: I do take some weekends off [laugh]. Sometimes when things need to be completed you have to find whatever time you have to do it!

FV: You are one of the most accomplished statisticians in activity, with lots of significant publications, you received the COPSS award eight years ago, and you are still in your mid-forties. What do you want to do with the rest of your life?

WW: I hope I can continue to be active and just keep enjoying scientific research! I think one is most lucky if he or she is

able to do what one really is interested in doing, so I consider myself very lucky. I think this is a wonderful environment to do that, so I will continue to do it. The other aspect is that the so-called achievements of mine are really team efforts. Almost all my papers are joint with various people, too numerous to name here, but it is just wonderful to have the chance to work with really talented graduate students and outstanding postdocs, and my colleagues at various universities have always been very supportive. I have all of those to be thankful for.

FV: Wing, I wish you to continue doing so and being an inspiration for my colleagues, myself, and the profession!

WW: Thank you very much!

## 2002 Biostatistics Schering-Plough Workshop

by Karen Kuntz

The 2002 Biostatistics Schering-Plough Workshop focused on emerging strategies in designing and monitoring trials, with speakers from academia, industry and government. The first session, chaired by Stephen Lagakos and Kenneth Koury, provided an overview of statistical approaches to clinical trial design and monitoring that allow for more flexibility, including a novel approach to chromosomal analysis. Presenters in this session included Ross Prentice (Fred Hutchinson Cancer Research Center), Janet Wittes (Statistics Collaborative, Inc.), Matthew Meyerson (Dana-Farber Cancer Institute) and Robert Temple (FDA). In the second session, chaired by Cyrus Mehta, James Ware (HSPH), Don Berry (MD Anderson Cancer Center), and Marvin Zelen (HSPH) discussed approaches for using Bayesian methods for bridging clinical trial data to new populations, making the drug development process more efficient, and formulating the sample size problem. Also in this session Peter Bauer (University of Vienna), Chris Jennison (University of Bath),

and Anastasios Tsiatis, (North Carolina State University) discussed approaches to adaptive designs for monitoring clinical trials, such as methodology to redesign an ongoing trial based on interim data. During the final session, chaired by Rebecca Betensky, the use and need for surrogate and predictive biomarkers of efficacy and toxicity was illustrated in applications by the Schering-Plough Working Group, Stephen Skates (Massachusetts General Hospital), and John Ryan (Wyeth Research), followed by a panel discussion with Jay Siegel (FDA), Thomas Haverty (Schering-Plough Research Institute), Gregory Campbell (FDA) and Stephen Lagakos. The symposium ended with the presentation of the 2002 Marvin Zelen Leadership Award to Robert O'Neill, Director of the Office of Biostatistics, Center for Drug Evaluation and Research, FDA. Dr. O'Neill provided a perspective on the development and future of statistics at the FDA. Presentations can be viewed online at [www.biostat.harvard.edu/events/schering-plough/agenda.html](http://www.biostat.harvard.edu/events/schering-plough/agenda.html).

## Biostat Students Graduate June 2001

In June 2001 the department awarded doctoral degrees to three students: Scarlett Bellamy, Jonathan French, and Knashawn Hodge Morales. Scarlett is now an assistant professor in the Department of Epidemiology and Biostatistics at the University of Pennsylvania. Jonathan had been working as a postdoctoral fellow in our department since he defended his thesis in September 2000 and began working at Pfizer, Inc. in Groton, CT at the end of July 2001. Knashawn accepted a position at New England Research Institute.

The department conferred Master's degrees to eight students: Carlos Brain, Hervé Caspard, Anna Choi, Karen Cormier, Shannon Harkness, Aparna Keshaviah, Yuhyun Park, and Xiaojing (Jean) Wang. Hervé Caspard and Anna Choi will continue working on their doctoral degrees in Epidemiology and Environmental Health, respectively, here at HSPH. Aparna joined the Department of Biostatistical Science at the Dana-Farber Cancer Institute beginning the end of July 2001. Jean is working at the Merck Research Lab, and Shannon is working at Boston Biostatistics Inc., a contract research organization located in Framingham. Yuhyun will be

continuing in the department working towards her doctoral degree.



Graduating Students from left to right: Knashawn Morales, Aparna Keshaviah, Jonathan French, Scarlett Bellamy, Xiaojing Wang, Carlos Brain.

## What We Are Up To These Days in Biostats!

by Stephen W. Lagakos, Department Chair

This has been another busy year for the Department of Biostatistics, highlighted by the activities of our faculty and students.

Joining the faculty for the coming academic year will be Tianxi Cai, Chengcheng Hu, Cheng Li, and Wei Wang. Tianxi (who graduated from our department in 2000) and Chengcheng will be based at HSPH and involved in our HIV/AIDS research through the Center for Biostatistics in AIDS Research (CBAR). Cheng and Wei will be based at DFCI and involved in cancer studies, with Cheng focusing on genomics. Also joining the department will be Molin Wang and Shirley Liu. Having just completed their doctorates, Molin and Shirley will be postdoctoral fellows during the coming academic year and become members of our faculty in July 2003.

We will miss the departure of Joe Ibrahim, Associate Professor, and Constantin Yiannoutsos, Senior Research Scientist in CBAR. Joe will be joining the Biostatistics Department at UNC as Full Professor and Constantin is going to Indiana University as Associate Professor and head of the biostatistics group at Indiana University's Cancer Center. We wish both well in their new and challenging positions.

In the fall of 2001, we had an outstanding and large (n=24) incoming class of students. While all shared a common interest in biostatistics and public health, this class was very diverse in background. In addition to broad representation from the U.S., our first year students hailed from such distant locations as Botswana, China, Mexico, Poland, Serbia, and Tanzania. These international students are complemented by a diverse group of U.S. students, making this talented class a great addition to our outstanding student body. This year's incoming class is also large (n=21) and a diverse and outstanding group.

Fifteen of our students received their degrees at commencement this year, including: Raji Balasubramanian, Jolene Birmingham, Bhaswati Ganguli, Erin Kammann, Steve Lake, Patricia Stephenson, and Lu Tian—who received their doctorates—and Jovanna Baptista, Nora Horick, and Ronnie Sebro, who received their Master's degrees and were present at the ceremony. As I sat on the podium awaiting the presentation of these students with their degrees, I marveled at what a talented and nice group of persons they are, and how fortunate we were to have had them in the department.

Beginning in the 2002-2003 academic year, HSPH will follow a new academic calendar, with the first term ending prior to the December holiday break. The second semester will be essentially unchanged, and the entire month of January will be a new "Wintersession" period in which short courses and other

educational activities will be offered. The most notable change to our curriculum is a result of the growing popularity of statistical genetics and genomics. Two (BIO227a and BIO228b) of the three courses in statistical genetics course that had been offered intermittently in the past were offered this year and, due to their popularity, will be offered again in the coming year, along with our more advanced course in this area (BIO268cd). We also will offer our new course in Computational Biology for the third straight year and, in addition, offer two other short genomics courses during the new Wintersession period.

Our partnerships with other universities and pharmaceutical companies have continued to provide us an excellent source of new ideas, experiences and opportunities. A new joint grant with the Karolinska Institute/University of Stockholm allows for exchanges of students and faculty. During the spring semester of 2002, we were joined by two Swedish doctoral students – Annica Dominicus and Maria Gruenwald. In the summer of 2002 one of our students – Yannis Jemai – is visiting the Karolinska to gain practical experience. Another of our students – Yu Guo – is spending the summer at Schering-Plough and learning about some cutting-edge problems in genomics, and over the past year four of our students – Karen Eckstein Han, Lu Tian, Mary Zhao, and Minhee Kang – have assisted our faculty at teaching assistants for 2-week courses we have offered at Kitasato University in Tokyo. We feel that these types of experiences provide valuable practical teaching/research experience for the students and also help them learn about how things are done in other parts of the world.

Our longstanding partnerships with the Genetics Institute (now part of Wyeth), Pfizer, and Schering-Plough have continued to generate valuable interactions and exposure to new challenges in study design and analysis; these also provide the department with much-needed support to finance the educational costs of our students. Two new partnerships have been formed this year, with Frontier Science & Technology Research Foundation and Eli Lilly Corporation. Frontier has been a generous source of support for student tuition and stipends in the past and our new partnership with them continues in the same spirit. Our partnership with Lilly is aimed at promoting community-based research, with a special interest in encouraging more students from minority US populations to seek careers in this area.

Finally, as in the past, so much of the success that our department has enjoyed is due to our faculty, who, in addition to their outstanding applied and theoretical research, continue to demonstrate great concern for the education and training of our students and the well being of the department.

## Neils Keiding Delivers 2001 Zelen Leadership Lecture

by Brent Coull

The Department of Biostatistics named Professor Niels Keiding, Professor of Biostatistics, University of Copenhagen the recipient of the 2001 Marvin Zelen Leadership Award in Statistical Science. The awarding ceremony took place on June 1, 2001 following the Schering-Plough workshop and a lecture by Professor Keiding on "Event Histories and Their Analysis". In his presentation, Professor Keiding surveyed his interests in event histories developing in calendar time, studied at a cross-section at a particular time.

This annual award, supported by colleagues, friends and family, was established in 1997 to honor Dr. Marvin Zelen's long and distinguished career as a statistician and his major role in shaping the field of biostatistics. The award recognizes an individual "in government, industry, or academia, who by virtue of his/her outstanding leadership, has greatly impacted the theory and practice of statistical science".

Professor Keiding exemplifies the criteria of the award. Throughout his career, he has worked to promote biostatistics as an academic discipline within the fields of Medicine and Public Health. He was a co-founder of the Danish Society for Theoretical Statistics in 1971. In 1978, he persuaded the Danish Research Councils to found a Statistical Research Unit within the Faculty of Medicine at the University of Copenhagen. Under Professor Keiding's guidance, this unit has grown into a formal Department of Biostatistics within the university.

Professor Keiding's leadership in biostatistics is apparent at an international level as well. He has served in an advisory role to promote medical statistics in Norway, Sweden, and Finland. He co-founded the Scandinavian Journal of Statistics and has served on its editorial board for many years. Beyond Scandinavia, Professor Keiding has served as president and

vice-president of the International Biometric Society, treasurer of the Bernoulli society, and vice president of the International Statistical Institute. He has served on the editorial boards of Biometrics, Statistics in Medicine and the Annals of Statistics, among others.

In addition to his impressive leadership, Professor Keiding has made important contributions in the areas of probability, theoretical statistics, and applied statistics. He is perhaps best-known for his work in the areas of counting processes and



From left to right: Dr. Richard Gelber, Dr. Niels Keiding, Dr. Marvin Zelen.

2001 Zelen

(cont'd on pg.8)

## Steven Goodman Delivers 2000 Lefkopoulou Lecture

by Brent Coull

Dr. Steven Goodman, Associate Professor of Oncology, Pediatrics, Epidemiology, and Biostatistics, John Hopkins University Schools of Medicine and Public Health, delivered the 2000 Myrto Lefkopoulou Distinguished Lecture at HSPH on September 21, 2000. His talk, "When Worlds Collide: Statistics, Ethics, and Clinical Research" focused on evidence-based statistics and its role in medicine, particularly randomized clinical trials. Dr. Goodman presented the viewpoint that the medical community often relies on the frequentist approach to applied research without a clear understanding of the long-standing debate among statisticians over the suitability of hypothesis tests and P-values for scientific inference. After tracing the history of this debate, Dr. Goodman presented the P-value fallacy, which states that one number can reflect both the long-run outcomes of an experiment and the evidential strength of a single result. He concluded the presentation by offering the Bayes Factor as an alternative measure of statistical inference that properly separates these two issues.



Dr. Steven Goodman

2000 Lefkopoulou

(cont'd on pg. 10)

## Bradley Carlin Gives 2001 Lefkopoulou Lecture

by Mei-Chiung Shih

Dr. Bradley P. Carlin, Professor in the Division of Biostatistics, University of Minnesota School of Public Health, was named the 2001 Myrto Lefkopoulou Distinguished Lecturer. Dr. Carlin presented a lecture on Thursday, September 20, at the Harvard School of Public Health. The title was "Hierarchical Models for Spatio-Temporally Correlated Public Health Data". In this lecture, Dr. Carlin surveyed recent developments in the analysis of spatial data, emphasizing spatial and spatio-temporal models appropriate for epidemiological and other public health datasets, and showed that a hierarchical framework, incorporating geographic information systems and Markov chain Monte Carlo methods, enables reasonably straightforward solutions to the analytic challenges in this area. A reception was held following the lecture.

Dr. Carlin obtained his Ph.D. in statistics from the University of Connecticut in 1989. He joined University of Minnesota since 1991. His research interests include statistical applications in AIDS research, spatial disease mapping, longitudinal studies, and the development of Bayes and empirical Bayes methods. Dr. Carlin is also the coauthor, with Dr. Tom Louise, of the textbook "Bayes and Empirical Bayes

Methods for Data Analysis", published by Chapman and Hall/CRC Press.

The lectureship was established in memory of Dr. Myrto Lefkopoulou, a faculty member and graduate of the department who died of cancer at the age of 34. Each year the lectureship is awarded to a promising statistician within 15 years of receiving an earned doctorate for contributions to collaborative or methodological research in statistical applications to biology or medicine, and/or for excellence in the teaching of biostatistics.



Dr. Brad Carlin

## 2001 Harvard/Schering-Plough Workshop

by Victor DeGruttola

The 2001 Schering Plough Workshop was entitled "Workshop on Datamining, with applications in genomics, clinical trials and post-marketing drug risk." A major focus of the workshop was on methods for analysis of high dimensional data; such as data arising from recent technological developments, like gene sequencing and gene expression arrays. As anyone involved in biomedical research is aware, these techniques have become increasingly common, but the best approaches for analysis is still a matter for discussion. The availability of such data have major implications for clinical research and practice, including the ability to target treatments to specific sub-populations defined by measures like genotype of microbes or human cells, surface antigens of cells, body burden of virus/tumors.

Workshop presenters came from academia, government and

industry, and their presentations ranged from a broad overview to analyses of specific data sets. In an example of the former, Daryl Pregibon from AT&T Labs described the current status and future prospects of the field of data mining, with a focus on the overlap of data mining methods with standard statistical techniques. Brad Efron from Stanford presented an example of the latter, in an analysis of data from a microarray experiment involving several thousand genes and several hundred thousand expression levels. In a similar vein, Rob Tibshirani from Stanford also discussed microarray data with a focus on class prediction. Stan Young from GlaxoSmithKline discussed methods borrowed from data mining for associating genes with disease phenotype, side effects and drug efficacy. There were also presentations from our Dept., including talks by Marcello Pagano, Andrea Foulkes and Victor De Gruttola.

## 2002 Zelen Award and Lecture

by Matt Wand

Dr. Robert O'Neill, Director of the Office of Biostatistics, Center for Drug Evaluation and Research, U.S. Food and Drug Administration (FDA), delivered the 2002 Marvin Zelen Leadership Award in Statistical Science on 31st May. It was titled "A Perspective on the Development and Future of Statistics at the FDA". Dr O'Neill described the laws and regulations impacting standards of evidence and pointed out that it creates a rich statistical environment He then

summarized evolutions of statistical program over 40 years. For the period 1962-1990 much energy at FDA went into establishment of basics and the approval process. Since about 1990 there has been markedly more pressure for faster approval, particularly fueled by the AIDS epidemic. Dr. O'Neill concluded with future areas of involvement and challenges.

## New Department Initiatives

by Marcia Testa

### Master of Public Health Degree Program in International Health for Physicians

Professor Marcello Pagano has been working with the School in its effort to establish a graduate degree in public health for physicians from less developed countries that are seeking a program that will be partly residential and partly at a distance. Professor Pagano has been strategically positioning the program to place a heavy emphasis on technology-based education by employing the internet and web-based learning. Tentatively, courses will be offered each year in a five-week summer session, followed by structured distance learning experiences, and a year-long field practicum. The degree will take two years to complete, including the two summer sessions. The objective of the program is to educate physicians who seek training in public health while they maintain their full-time jobs. The program targets mid-career physicians who have completed their clinical training and have practiced for at least three years in a specialty or in public health.

### Summer Only Master of Public Health in Quantitative Methods

Marcia A. Testa is currently the advisor for the 20 students currently enrolled in the School's new summer-only option for the Master of Public Health (MPH) degree program in Quantitative Methods. The program jointly sponsored by the Departments of Biostatistics and Epidemiology is completed over the three summers and is available to professionals who have doctoral or masters training in fields such as medicine, dentistry, veterinary medicine, law. Students are based throughout the United States and internationally in countries such as Canada, Sweden, Great Britain, Australia, Japan and Germany. In-between summers, the students work at their home-based institution on an independent research practicum which they submit and present during their final summer. The

group communicates and keeps up to date during the year using the School's Web-Based Blackboard system, a comprehensive and flexible e-Learning software platform. The first graduate of the program Ed Smith, M.D., M.P.H., received his degree in June 2001.

### CBAR expands HIV/AIDS research collaboration with the Government of Botswana

In 1996, the Harvard AIDS Institute (HAI) and the Government of Botswana initiated a long-term research collaboration concerning HIV/AIDS. During the past year, HAI's activities have increased greatly in scope and several CBAR and Department members have begun providing much-needed statistical expertise. A large randomized clinical trial (the "Mashi" or "Milk" trial), designed to assess the efficacy of different antiretroviral regimens and infant-feeding strategies for the reduction of mother-to-child transmission, opened to accrual in April. Peter Gilbert, Laura Smeaton, and Steve Lagakos collaborated in the design of this trial. Peter Gilbert and Victor DeGruttola collaborated in the design of a second randomized clinical trial (the "Tshepo" or "Hope" trial), which will open soon, to compare the potency, tolerability, and development of drug resistance in adults treated with different antiretroviral combination therapies and to assess the efficacy of directly-observed therapy. David Shapiro and Ron Bosch collaborated in the design of a third randomized trial, submitted for funding, which proposes to study triple combination antiretroviral therapy in pregnant women to reduce mother-to-child transmission and different treatment strategies for maternal HIV disease (directly-observed therapy and structured-treatment interruption). Michael Hughes is collaborating in a proposal for long-term follow-up of mothers and babies participating in the Mashi study.

*Department Initiatives*

*(cont'd on p. 11)*

## Biostat Students Graduate June 2002

This year the department awarded doctoral degrees to eight students: Raji Balasubramanian, Jolene Birmingham, Bhaswati Ganguli, Erin Kammann, Steve Lake, Patricia Stephenson, Max Su, and Lu Tian. Raji is now working as a Research Associate in the Center for Biostatistics in AIDS Research. Jolene is taking some time off with her family before job hunting. Erin is a Senior Biostatistician with the Department of Medical Affairs at Alkermes, Inc. in Cambridge, MA. Pat and Lu will both be research fellows in our department in the Fall 2002. Steve and Erin just got married, and Steve will be job hunting this summer.

The department conferred Master's degrees to seven students: Jovanna Baptista, Nora Horick, James Hudson, Ronnie Sebro, Maria Shubina, Chang-Heok Soh, and Clarissa Valim. Dr. James Hudson will continue his work as a professor in Biological Psychiatry Laboratory at McLean Hospital in Belmont, MA. Clarissa will continue working on her doctoral degrees in the Department of Infectious Diseases here at HSPH. Ronnie, Chang-Heok and Maria will be continuing in our department working towards their doctoral degrees.



**Graduating Students from left to right: Erin Kammann, Stephen Lake, Jolene Birmingham, Lu Tian, Bhaswati Ganguli, Raji Balasubramanian, Patricia Stephenson.**

## News and Updates

by Rebecca Betensky

This has been a busy and productive year! Here is a listing of some of our news:

- In Fall, 2001, we welcomed 25 students into the Department. Of these, 21 enrolled in our doctoral program, 1 in our master's program, 1 is a special student, and 1 is a Fogarty Fellow. Of the degree students, 19 are US citizens or Permanent Residents, and 4 are international students. Our Fogarty Fellow is from Botswana.
- Professors James Robins and Andrea Rotnitzky offered a mini-course titled "Semiparametric Theory, Missing Data, and Causal Inference Problems."
- We are pleased to announce that two of our faculty received Committee of Presidents of Statistical Society (COPPS) awards at the 2001 Joint Statistical Meetings. Professor Nan Laird received the first F.N. David Award for leadership and serving as a role model among women in the profession. Professor James Robins received the George Snedecor Award for his publication "Semiparametric efficiency and its implications on the design and analysis of group-sequential studies" with Professors D.O. Scharfstein and A.A. Tsiatis.
- The department offered a short course in January titled "Applications of Statistics to Genomics." The course covered the analysis of gene expression data from DNA microarray experiments and nucleotide and protein sequence analysis. The instructors were Professors Robert Gentleman, Sandrine Dudoit (U.C. Berkeley), Vince Carey, Lior Pachter (U.C. Berkeley), and Cheng Li.
- We congratulate Helena Kraemer, Professor of Biostatistics in Psychiatry at Stanford University, on receiving the Harvard Award in Pyschiatric Epidemiology and Biostatistics. The title of her award lecture was "Statistics, Language, and Medical Science."
- The Department welcomed Annica Dominicus and Maria Grunewald in the Spring 2002 semester.

Annica and Maria are doctoral students at the University of Stockholm.

- We are pleased to announce that Professor Garrett Fitzmaurice was selected as a recipient of an HSPH Junior Faculty Sabbatical Award for the 2001-2002 academic year. He began work on a book on methods for analyzing longitudinal data. Professor Florin Vaida was selected as a recipient of an HSPH Junior Faculty Sabbatical Award for the 2002-2003 academic year. He plans to use this time to extend his research interests in HIV/AIDS and genetics.
- We congratulate Professor Garrett Fitzmaurice on his election as a member of the International Statistical Institute.
- We congratulate Professor Yi Li on being awarded the Janssen Research Foundation (Belgium) Chair in Survival Analysis for the 2002-2003 academic year. He will teach survival analysis at Limburgs Universitair Centrum, Belgium in 2003.
- We congratulate Professor Richard Gelber, who has been named as the recipient of the European Institute of Oncology Breast Cancer Award for 2002. Rich will be presented the award and a golden medal in a ceremony on June 5, 2002, at the 4<sup>th</sup> Milan Breast Cancer Conference.
- We congratulate Professors Matt Wand, Michael Hughes and Sharon-Lise Normand on their election to Fellows of the American Statistical Association.
- We congratulate Jared Christensen on receiving a Teaching Assistant Award, Professor Kimberly Gauvreau on receiving a Teaching Award, and Professor Matt Wand on receiving a Mentoring Award.
- We are pleased to announce the promotion of Professor Sandra Lee to Senior Research Scientist.

2001 Zelen

(cont'd from p. 5)

failure time data analysis, as demonstrated by the textbook "Statistical Models Based on Counting Processes" with co-authors Andersen, Borgan, and Gill.

Professor Keiding was born in Copenhagen in 1944. After receiving his M.Sc. degree in Statistics from the University of Copenhagen in 1968, Professor Keiding served as an Assistant and Associate Professor in this department until 1978. He then served as Research Director of the Statistical Research Unit for six years. In 1984, he became Research professor in

Biostatistics on the Faculty of Medical Sciences at the University of Copenhagen, and a Professor of Biostatistics in this department in 1990. Since 1992, Professor Keiding has also served as an Adjunct Professor in the Department of Statistics at Ohio State University.

Professor Keiding is the fifth recipient of the Zelen Leadership Award, which was previously conferred on Frederick Mosteller, Sir David Cox, John Tukey, and Lincoln Moses.

## An Interview with Graduating Students

by Bhaswati Ganguli and Nora Horick

The department said goodbye to several students over the past two years. These include Aparna Keshaviah who was awarded a Master's degrees, and Scarlett Bellamy, Jonathan French and Jolene Birmingham, who completed doctoral degrees in 2000-2001. Last spring we asked each to offer his or her words of wisdom regarding both past and future experiences. We also interviewed Erin Kammann, Stephen Lake, and Raji Balasubramaniam who have just graduated. We present below, a selection of responses:

What are you doing now?

- Aparna: I will begin working at Dana-Farber July 30th.  
 Scarlett: I'm starting as Assistant Prof at University of Pennsylvania on Sept 1.  
 Erin: I have accepted a position at Alkermes, a Massachusetts based biotech/pharmaceutical company specialising in drug delivery. Alkermes has an academic feel and publication goals and my focus will be largely development of methodology as well as applications.  
 Jonathan: I'm currently doing a postdoc in the department with Louise Ryan. In mid-July I will be starting a job with Pfizer, Inc.  
 Steve: Attempting to finish my thesis work and riding my bike.  
 Raji: I am finishing up my thesis.  
 Jolene: Writing the introduction to my thesis and preparing slides for my defense. I am taking a sabbatical after I defend.

How has your perspective on the field of Biostatistics changed since you started the program?

- Scarlett: I think I have a better 'big picture' view.  
 Erin: I have a much deeper understanding of what's entailed in statistical methodological research and how biostatistics is used across so many disciplines.  
 Steve: I have realized that the field of biostatistics encompasses much more than the standard clinical trials. There are a wide range of potential career lines suitable to each individual's talent.  
 Jolene: The theoretical aspects to statistical methodology are much more interesting to me than before I began the program. It is exciting to be in a department which undertakes such a variety of research topics!

What are the areas (if any) that you wish you had received more exposure to while you were a student?

- \*: Study design.  
 Aparna: Perhaps some applications of biostatistics to genetics.  
 Scarlett: Just more practical experience to real projects; possibly even some exposure to grant writing, etc.  
 Erin: Applications. In my experience here, the main place to actually practice statistics as a student is to

TA the good applied courses such as rates and proportions, survival, discrete data analysis and longitudinal analysis for instance. The other way is to get involved in some projects outside of your thesis research. Students simply don't have enough time to take all the courses which would completely round out the experience, so it's imperative that the department encourages them to TA the applied classes in order to get that exposure.

- Jonathan: Applied work and medium-term collaborative projects (i.e., projects that may last for one or two semesters, in addition to the type of projects that we get in the consulting course).  
 Raji: I would have liked to have more exposure to applied/methods courses, which would particularly be helpful in industry.  
 Jolene: Applications to real datasets, including exploratory data analysis, diagnostic tools and graphical techniques to assess model fit.

What was the job search process like?

- \*: Slow at first but once interviews started, they all came in at once. That led to a lot of juggling between classes, homework, and interviews.  
 Aparna: It was actually very easy for me because I started early (in October!). The hardest parts were getting my CV up to speed and figuring out where I wanted to apply.  
 Scarlett: A bit overwhelming... kinda like a 15 credit course! There are talks to prepare (if you are asked to give a presentation) which may differ greatly from place to place (e.g., time limit, content, etc.); travel plans which includes taking off time from thesis research and/or classes; administrative things to do (prepare CV, cover letters, etc); respond immediately to inquiries from prospective employers etc., etc.  
 Erin: Very rewarding! I applied for jobs both in industry and academia and offers abounded. The hardest choice for me was between academic and industry-type positions, and of course, location (which depended heavily on one fellow student in the department!). The compensation packages are excellent in both fields and I encourage students to always negotiate salaries and benefits, vacation time etc. All employers expect it now.  
 Jonathan: It was both fun and exhausting. The fun part: I really enjoyed meeting other statisticians in their work environment. The exhausting part: most interviews last for two days.  
 Steve: Stressful but informative. For those not looking into academics I would advise them to not worry about interviewing until they are just about done, unless they see an advertisement that really strikes their fancy. Of course, academics requires interviews typically 9 months prior to the proposed start date so plan accordingly.

*Student Interview*

*(cont'd from prev page)*

What qualities were potential employers looking for? Were they interested in course work/research or applied experience?

- \*: One of the two following combinations is desirable: SAS / clinical biostats or genetics / Matlab / C / Splus / R.
- Aparna: Some employers were impressed that I had already published an article, so that seemed to be important to the (academic) places to which I applied.
- Scarlett: For me, since I think it was more 'potential', competence, ability to collaborate/communicate with others (statisticians, MDs, epidemiologists, etc.).
- Erin: I would have to say both. It really depends on the position and the employer. Everywhere I interviewed, though, acadmia and industry, employers were simply looking for bright people with solid statistical and communication skills. Of course, a good publications list and some prior experience is also very helpful.
- Jonathan: Most potential employers were looking for someone who was a good statistician and who could communicate well with non-statisticians; the expectations for theoretical work depended on the job.
- Steve: It depends on what job you are trying to get. The standard pharmaceutical jobs just want someone who is bright and who has shown they can do competent research. Academics are always trying to fill what they consider to be their weak point and this usually happens to coincide with what is the hottest area in the field.

What advice would you offer the newer students in the program?

- \*: Take as many basic science classes (physiology, immunology, genetics, etc.) as you can. Don't worry if you get a bad grade. This is the most important knowledge you will rely on when interacting with scientists across disciplines.
- Aparna: I would tell them to work hard, to not get caught up in grades and the competition between students, and to remember to have fun!
- Erin: I would say try to take a breadth of coursework areas, and definitely TA at least 1 or 2 of the classes I mentioned above. Then, don't take courses for too long -- The amount you learn doing thesis work will be amazing and the learning always continues long after we leave here. Work really hard, and then play hard (windsurf or play frisbee!).
- Jonathan: Get as much practical experience as you can. It can help you decide on the type of job you would like and to get that job.
- Steve: The most important thing is to get the degree. Do not waste time worrying about the applicability or relevance of your work and do not get sidetracked with extraneous projects.
- Raji: Take time to take as many classes as you can and make use of the university resources outside of the biostatistics department.
- Jolene: Begin reading articles from statistical journals as soon as you begin gaining an interest in a topic. They will give you insight into the material required for a thesis paper and an idea of what types of statistical problems are solved within the boundaries of a research paper.

\* These responses are from three other graduating students who wished to remain anonymous.

*2000 Lefkopoulou*

*(cont'd from pg. 5)*

Dr. Goodman received his MD from New York University and did his residency in pediatrics at St. Louis Children's hospital before earning a PhD in epidemiology from Johns Hopkins in 1989. While a student in the Department of Epidemiology, he earned a joint master's degree in Biostatistics and ultimately wrote his PhD thesis under the direction of Richard Royall in Biostatistics.

The Myrto Lefkopoulou Lectureship was established in perpetuity in memory of Dr. Myrto Lefkopoulou, a faculty member and graduate of the department. Dr. Lefkopoulou tragically died of cancer in 1992 at the age of 34 after a

courageous two-year battle. Each year the lectureship is awarded to a promising statistician who has made contributions to either collaborative or methodologic research in the applications of statistical methods to biology or medicine, or has shown excellence in the teaching of biostatistics. Ordinarily, the lectureship is given to a statistician who has earned a doctorate in the last fifteen years.

Nominations for the 2003 lectureship are welcome and should be sent to the department or made electronically at <http://www.biostat.harvard.edu/events/awards/myrto/> by March 15, 2003.

*Department Initiatives*

*(cont'd from p. 7)*

**The Microarray Database and Analysis Project**

Robert Gentleman, Wing Wong, Vince Carey, Peter Park, Byron Ellis, Ming-Chih Kao, Cheng Li are working on a project to develop an integrated environment for the analysis and storage of microarray data. This project will produce high quality, architecture neutral software for the collection, storage and analysis of data arising from experiments that involve DNA microarrays. The project will be both open source (the code is available to all) and open development (others, with suitable skills, may contribute). The "system" will comprise a data warehouse together with translation and analysis tools. The data warehouse will have two components. Large data components (e.g., image data) will be stored in HDF5 format. Other data, typically patient level covariates, will be stored in a relational database. There will be strong links, via software between the two components. The translation tools will mainly consist of a set of scripts in various languages (such as C/C++, Perl, Python) that will translate the image data into various formats suitable for analysis. The analysis tools will be mainly based in R ([www.r-project.org](http://www.r-project.org)), and implementation of the ACM award winning S language. These tools will provide state of the art implementations of current and future tools for analysing microarray data.

**Modeling Treatment Use & Effectiveness in Mental Illness**

Sharon-Lise Normand, Mary Beth Landrum, and Philip Lavori in Biostatistics joined with faculty members throughout the University including Richard Frank (Professor of Health

Economics), Marcela Horvitz-Lennon (Instructor in Psychiatry), Haiden Huskamp (Assistant Professor of Health Economics), Mary Beth Landrum (Assistant Professor of Biostatistics), Thomas McGuire (Professor of Economics), Donald Rubin (Professor of Statistics), Joel Greenhouse (Professor of Statistics), and Robert Rosenheck (Professor of Psychiatry) to examine the development and application of discrete choice models for understanding treatment use and for causal inferences in experimental and naturalistic studies of mental illness. By studying how patients are matched with treatments in extant systems, the researchers involved will gain greater insight into the determinants of quality of care. The research will provide for the 1) extension of likelihood-based methods to estimate treatment effectiveness at the levels actually received using experimental data from two influential clinical trials and to compare these estimates with those based on conventional approaches, such as intention-to-treat, adequate, and completer principles, 2) development of new models of discrete choice to explain variation in treatment use based on patient, provider, and insurance characteristics for privately insured and Medicaid beneficiaries, and 3) application of these discrete choice models to explain variation in adherence with treatment recommendations and in treatment effectiveness for depression and for schizophrenia across a diverse array of practice settings. The methodological advances from this federally funded research will enable mental health researchers and policy makers to better characterize usual care and to expand the inferences drawn from clinical trials.

**New Research Fellows in 2001-2002**

This past year the department welcomed five new research fellows who bring with them a variety of research interests and come from diverse backgrounds.

**Long Ngo** is from the San Francisco Bay Area. He received his degree from UC Berkeley, in Biostatistics. His research interests are in mixed models, model selection, and recently smoothing methods using mixed models. He is currently working with Dr. Matt Wand, and Prof. Ryan. He arrived here in December of 2001, and so far he really likes Boston. His hobbies are windsurfing, swimming, table tennis, guitar, reading, and ballroom dance, especially Argentine Tango.

**Haiyan Huang** was born in Shenyang, a city located at the north of China. She got her Bachelor degree of Mathematics from Peking University, China in 1997. In the same year, she came to USA to pursue my Ph.D. During the four years in the University of Southern California, Ms. Huang concentrated on the study of probability, statistics and biology. In 2001, she completed her Ph.D. dissertation with Prof. Larry Goldstein in the Center for Computational and Experimental Genomics. Currently, Dr. Huang is a post-doctoral research fellow in the department of Biostatistics, Harvard School of Public Health. Haiyan is working with Prof. Wing H. Wong and Prof. Jun S. Liu on developing novel methods to identify transcription factor binding sites in the regulatory regions in the genome. Her research interests include the application/development of statistical methods to biological sequences, and biological knowledge discovery through the analysis of SAGE data, etc.

**Nandita Mitra** is a post-doctoral research fellow in Statistical Genetics. She is working with Nan Laird on family-based genetic association methods. She is partially funded by a National Institute of Mental Health post-doctoral training grant. Nandita received her BA from Brown University, her MA from UC Berkeley and her Ph.D. from Columbia University. She recently accepted a faculty position in the Department of Biostatistics at Memorial Sloan-Kettering Cancer Center with a joint appointment at Weill Medical College of Cornell University. She will be moving to NY in August. Nandita had a great time exploring Boston (especially the charming bookstores) this past year and is sad to leave friends and colleagues at Harvard.

**Andrew Strahs** currently works with Wing Wong on problems in computational biology and bioinformatics. In 2001, he received his Ph.D. in Statistics from the Department of Statistics at the University of Chicago. His dissertation, supervised by Mary Sara McPeck, discussed their model-based approach to the fine-scale mapping problem in statistical genetics. He lives in Cambridge with his wife, Colleen, and their fledgling herb garden (not yet named).

**Anthony Hamlett** received his Ph.D. from the University of Rhode Island in Applied Mathematical Sciences in 1999. His current interests are in repeated measures, dose allocation and environmental health. He is currently working with Drs. Louise Ryan and Matt Wand. He enjoys playing soccer.

## New Research Fellows in 2000-2001

Our department continues to be enriched by postdoctoral researchers with a variety of backgrounds and research interests. Last year we were fortunate to welcome seven new postdoctoral fellows.

**Yan Cui** is originally from China and received his Ph.D. in Biophysics from the Institute of Biophysics, Chinese Academy of Sciences in 1998. He did postdoctoral training with Dr. Wing Wong at UCLA before coming here to Harvard to continue work with Dr. Wong in his Bioinformatics Lab. His research interest include biological knowledge discovery through the analysis of microarray data, protein structure prediction and computer simulation, computer simulation of protein evolution, evolutionary theories of sex and aging, and modeling biological system, such as genetic network and signaling networks of the cell.

**Hongyu Jiang** is also originally from China and received her Ph.D. from the Department of Statistics at the University of Wisconsin-Madison. Hongyu arrived as a research fellow in August, 2000 and became an assistant professor on July 1. Hongyu is excited to be working in our department and collaborating with CBAR. She enjoys doing research in multivariate survival analysis with semi-parametric or non-parametric methods. Her research interests are in the area of innovative design and analysis of clinical trials. Moving from the midwest to Boston has been a big change for Hongyu, but she appreciates the charm of this city more than her complaints about the traffic. In her spare time, Hongyu likes playing tennis, hiking and reading non-statistics books.

**Christoph Lange** finished his PhD in Statistical Genetics at the University of Reading in the summer of 2000 and started his postdoctoral fellowship here last September. Before that he received a "Diplom" in Mathematics and Computer Science in Germany and an MSc in Biostatistics in Belgium. Christoph's research interests are statistical genetics and generalized linear models. He spends his spare time playing volleyball and tennis, listening to classical music, sailing and eating lots of chocolate.

**Igor Leykin** has both an MD degree and a PhD degree in biology. For the last 8 years he has been involved in research related to immunology and diseases of the immune system. He is interested in studying new approaches which have become possible due to recent discoveries in genomics. Igor sees this field as very promising for future practical applications in medicine.

**Ovidiu Lipan's** background is in theoretical physics. He studied topics related to the quantization of solitons and integrable systems in quantum field theories at the University of Chicago. After Chicago, he moved to sunny Pasadena, where he was a Sherman Fairchild Postdoctoral Fellow in the Division of Physics Mathematics and Astronomy at California Institute of Technology. There he worked on a mathematical problem connected with the motion of electrons in two dimensions. During his stay at Caltech he realized that biology is on the verge of becoming a quantitative science. At that time he met Professor Wing Wong's group at UCLA. At the end of April 2000, he decided to join the group and as a consequence Ovidiu, together with his family, moved to the east coast. Ovidiu likes to visit sites around Boston, in the range of a radius of 300 miles, and to learn about climate differences between regions.

**Peter Park** came to the department as a student in the fall of 1999, having a PhD in applied mathematics (numerical analysis for partial differential equations). After a year of classes, he became a postdoctoral fellow, thus successfully avoiding the prospect of failing the qualifying exam. Since then, he has been working with Drs. Pagano and Bonetti on gene expression data. Starting this fall, he will be an Instructor as part of the Informatics Program at the Children's Hospital, continuing to work on microarray data analysis and functional genomics in general. Although he will not be officially affiliated with the department, he will be continuing his collaborations with various people in the department.

**Xianghong Zhou** received her undergraduate degree in biochemistry from the University of Tuebingen, Germany. From 1997 to 2000, she pursued her Ph.D degree at the Swiss Federal Institute of Technology (ETH). The subject of her doctoral study was bioinformatics with a focus on the protein structure and proteomics. Before joining HSPH, during the fall semester of 2000 she was a postdoctoral fellow at the UCLA within the special program "Functional Genomics". Xianghong is interested in many topics in bioinformatics including microarray analysis, mass pectrometry and sequence analysis.