

BIOSTATISTICS

JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

Annual Retreat

April 29 - May 1, 2005

THE HOTEL HERSHEY
P.O. Box 400
Hotel Road
Hershey, PA 17033-0400
Telephone: 717-533-2171
Front Desk FAX: 717-534-8887

BIostatistics RETREAT AGENDA
THE HOTEL HERSHEY
APRIL 29TH, 2005 – MAY 1ST, 2005

Friday, April 29th:

- 4:00-6:00 PM: Student & Faculty Scientific Poster Presentations Happy Hour
(Mezzanine Room)
- 6:00-8:00 PM: Asian Escape Dinner (Castilian Room)

Saturday, May 1st:

- 7:00-10:00 AM: Breakfast (Circular Dining Room)
- 9:00-10:00 AM: Background Information for Morning Session - Scott Zeger
(Garden Terrace West)
- 10:00-10:15 AM: Break
- 10:15-11:30 AM: Break Out Session (Tea House, Rose Garden, Cocoa Inn Room)
- 11:30-12:30 PM: Reports from groups and summary of findings (Garden Terrace West)
- 12:30-2:00 PM: Lunch-Jolly Rancher (Fountain Lobby)
- 2:00-2:45 PM: Scientific Presentation and Discussion - Section 1 (Garden Terrace West)
Dominici, Scharfstein, Tan and Zeger
- 2:45-3:00 PM: Break
- 3:00-3:45 PM: Scientific Presentation and Discussion - Section 2 (Garden Terrace West)
Bandeem-Roche, Crainiceanu, Parmigiani
- 3:45-4:00 PM: Break
- 4:00-4:45 PM: Scientific Presentation and Discussion - Section 3 (Garden Terrace West)
Caffo, Louis, Ruczinski, Yin and Irizarry
- 4:45-6:00 PM: Free Time
- 6:00-8:00 PM: Carnival de Hotel Hershey Dinner (Garden Terrace West)
- 8:00-9:30 PM: Movie- Willie Wonka & The Chocolate Factory (Garden Terrace West)

**DEPARTMENT OF BIostatISTICS
TABLE OF CONTENTS
2005**

2	Administrative Information
8	Qualitative Survey of Public Health Scientists and Professionals
10	Summary of Responses to Qualitative Survey
12	Responses to Qualitative Survey
23	Breakout Group List
24	Questions to Address
25	Student and Faculty Poster Presentation Session
27	Biostatistics Retreat Scientific Presentations

**State of the Department
April 2005**

Table 1: Number of Full-Time Faculty, Students, and Staff

	FY01	FY02	FY03	FY04	FY05
Faculty					
Full Professor	8	8	9	9	9
Associate Professor	1	1	3	3	5
Assistant Professor	5	7	4	6	4
Total tenure-track	14	16	16	18	18
Instructor	0	0	1	2	1
Research Associate	3	3	3	3	3
Scientist	4	3	3	3	3
Total Non-tenure-track	7	6	7	8	7
Total Faculty	21	22	23	26	25
Staff	6	6	7	8	8
Postdocs	1	1	2	2	2
Biostat students					
PhD	29	30	37	50	43
Master's	8	7	6	2	5
MHS	2	1	0	0	2
ScM	6	6	0	2	3
Total	37	37	43	54	47
Courses	54	57	57	60	59
Enrollments	2785	2777	2877	2915	2917

**State of the Department
April 2005**

Table 2: Department of Biostatistics Revenues in US Dollars (\$1,000).

Source	FY 2001	FY 2002	FY 2003	FY 2004	Projected FY 2005
General Funds (GF)					
TAM	941	1,080	1,233	1,311	1641
F&A	422	551	544	674	645
Total GF	1,363	1,631	1,777	1,985	2286
Sponsored Projects*	1,902	1,637	2,103	2,203	1,758
Total Direct					
Outside salary support **	936	1,062	1,228	1,532	1,547
Computer Services (BCSS)	127	141	149	199	120
Consulting Center	496	272	332	236	296
Total Operating Budget	4,142	4,291	4,663	6,066	6,007
Endowment Market Value	2,333	3,870	4,019	5,401	5,800

- *Biostatistics PI, Total Expenses Direct Expenses from CICS
- ** Non-Biostatistics PI, includes salary and fringe from Biostatistics Salary Spreadsheet.

**State of the Department
April, 2005**

Table 3: Student Data for the Department of Biostatistics, 1995-2005

	95-96	96-97	97-98	98-99	99-00	00-01	01-02	02-03	03-04	04-05
Applicants*	87	87	79	92	94	102	126	163	224	187
Accepted and Funded*	N/A	N/A	6	8	9	14	12	13	16	19
Enrolled**(new only)	12	10	9	7	8	12	10	15	16	16
Doctoral	3	6	4	2	4	9	7	12	12	12
(Funded doctoral)	?	(4)	(4)	(2)	(3)	(9)	(6)	(7)	(5)	(11)
Master's***	6	4	4	5	3	2	3	2	2	3
PDFs	3	0	1	0	1	1	0	1	2	1
Courses Offered	41	41	41	49	55	54	57	57	60	59
Baltimore	33	32	33	42	43	38	38	38	42	42
Montgomery County	6	6	5	5	5	5	4	4	3	2
Summer & Winter Institutes	2	3	3	2	5	6	10	11	11	13
Distance Ed	N/A	N/A	N/A	0	2	5	5	4	4	2
Enrollments	2102	2340	2018	2382	2603	2785	2777	2877	2915	2917 ~
Baltimore	1915	2148	1844	2199	2342	2451	2341	2408	2498	2547 ~
Montgomery County	158	161	148	172	163	130	126	95	37	39 ~
Summer & Winter Institutes	29	31	26	16	39	67	108	160	157	159 ~
Distance Ed	N/A	N/A	N/A	0	59	137	202	214	223	172 ~
Credits Earned	8533	9495	8220	8362	8774	9588	9307	9870	10643	10753 ~
Baltimore	7728	8676	7471	7724	8029	8655	8139	8576	9502	9747 ~
Montgomery County	665	670	622	562	544	406	404	318	148	117 ~
Summer & Winter Institutes	140	149	127	76	83	172	247	359	342	373 ~
Distance Ed	N/A	N/A	N/A	0	118	355	517	617	651	516 ~

* Does not include postdoctoral fellow or special student applications

** Does not include special students

*** Does not include joint MHS-PhD students

~ Projected

NOTES:

Data on applicants, accepted, enrolled from departmental files

Course and enrollment data are from the Registrar's Office's course enrollment reports; data excludes all 140.8— (special studies, thesis research, MPH Capstone) registrations; includes all interdivisionals; credits earned by Homewood students converted to PH credits.

**State of the Department
April, 2005**

Table 4: Recent Biostatistics PhD Graduates
Academic Years 2003-2004 and 2004-2005

Name	Academic Year Graduated	Advisor	Thesis Title	Current Position	Academic Years to Complete PhD
Leena Choi	04-05	Caffo	<i>Modeling Biomedical Data and the Foundations of Bioequivalence</i>	Asst Prof Vanderbilt U Dept of Biostatistics	6
Michael Griswold	04-05	Zeger	<i>Complex Distributions, Hmmm... Hierarchical Mixtures of Marginalized Multilevel Models</i>	President Griswold Consulting	7
Dongmei Liu	04-05	Parmigiani	<i>Application of Hierarchical Models in Microarray Data Analysis: Screening for Differentially Expressed Genes and Making Inference on Functional Classes</i>	Research Fellow London School of Hygiene & Tropical Medicine Dept of Infectious & Tropical Diseases	5
John Robinson	04-05	Zeger	<i>A Hierarchical Multivariate Two-Part Model for Profiling Providers' Effects on Healthcare Charges</i>	President John W. Robinson, MD, PhD, LLC	9
Michelle Shardell	04-05	Scharfstein	<i>The Analysis of Informatively Coarsened Discrete Time-to-Event Data</i>	Asst Prof U of MD Dept of Epi & Preventive Medicine	5
Ravi Varadhan	04-05	Frangakis	<i>The Role of the Design, Analysis, and Computation in Addressing Aetiology in Three Types of Studies in Public Health</i>	Asst Prof JHU Dept of Medicine	6
Zhijin Wu	04-05	Irizarry	<i>Probe Level Models for DNA Microarrays</i>	Asst Prof Brown U Ctr of Statistical Sciences	5
Weimin Chen	03-04	Broman	<i>Robust Quantitative Trait Linkage Analysis in Extended Human Pedigrees</i>	Postdoctoral Fellow U of Michigan Ctr for Statistical Genetics	4
Wesley Eddings	03-04	Rohde	<i>Topics in the Philosophy of Statistics: Methods, Data, and Theory</i>	Asst Prof Birmingham-Southern College Div of Sci & Math	6
Nikhil Gupte	03-04	Brookmeyer	<i>Statistical Models and Methods for Mother to Infant HIV Transmission Studies</i>	Data Manager/Statistician Johns Hopkins Department of Medicine, Division of Infectious Diseases	5

**State of the Department
April 2005**

Table 5: Support for Full-time PhD Students with Departmental Funding
(in thousands of dollars)

	FY01		FY02		FY03		FY04		FY05*	
Number of students ever enrolled in year	26		29		29		37		47	
Tuition										
Department	\$425	77%	\$551	80%	\$554	78%	\$711	89%	\$1,012	87%
Grants	\$126	23%	\$145	20%	\$158	22%	\$89	11%	\$148	13%
Total	\$552	100%	\$696	100%	\$712	100%	\$836	100%	\$1,160	100%
Health Insurance										
Department	\$13	59%	\$17	57%	\$23	63%	\$39	80%	\$44	85%
Grants	\$9	41%	\$12	43%	\$13	37%	\$10	20%	\$8	15%
Total	\$23	100%	\$29	100%	\$36	100%	\$49	100%	\$52	100%
Stipend/Wages										
Department	\$45	11%	\$50	11%	\$124	18%	\$54	9%	\$216	31%
Grants	\$352	89%	\$415	89%	\$573	82%	\$574	91%	\$476	69%
Total	\$398	100%	\$465	100%	\$697	100%	\$628	100%	\$692	100%
Total Support	\$972		\$1,190		\$1,441		\$1,513		\$1,904	

- Projection based on the following:
 - Tuition from the DGA Report
 - Health Insurance from the CICS System
 - Stipend/Wages from the CICS System – Department numbers include General Funds, Biostatistics Center and Gift Accounts.

**State of the Department
April 2005**

Table 6: Johns Hopkins Biostatistics Center:
Revenue Summary by Client Category in US Dollars (\$1,000)

	FY01	FY02	FY03	FY04	FY05*
JHMI	95	83	88	131	141
External	400	188	244	105	155
Total	495	272	332	236	296

*Projection annualized based on February 2005 data.

**DEPARTMENT OF BIOSTATISTICS
QUALITATIVE SURVEY OF PUBLIC HEALTH SCIENTISTS AND PROFESSIONALS
2005**

This qualitative survey seeks your opinion about the 3-5 year future of research opportunities in your area of expertise and the possible needs for biostatistical expertise. We very much appreciate your answering the questions below and providing any other thoughts in the space provided at the end.

If some questions are not appropriate, please just indicate N/A and answer the others.

If it would be easier for you, we would be happy to arrange a phone call to solicit your responses orally. Call Stephanie Panichello at 410-955-3067 or email to spaniche@jhsph.edu.

We will use your ideas and those of others as input to our 2005 retreat. We will prepare a document summarizing the collective thinking of the survey respondents and department and share it with you in a few months.

Thanks very much in advance sharing your ideas by April 22, 2005.

Name _____

Department _____

1. What are the open scientific questions that currently drive research in your field?

2. What changes have occurred in the way the best studies in your field are currently being done and what has caused them?

3. What new measurement techniques are having a major impact on how research is conducted in your field?

4. What new quantitative methods are increasingly popular in your field?

5. What quantitative methods does your group seek to develop further expertise in during the coming period?

6. How can the Department of Biostatistics be more supportive of your research program?

7. What paper should we read to better understand the important trends in your area of research?

8. What else would you like to tell us to make us a better department and/or more useful to your group?

Many thanks for taking the time to complete this questionnaire.

Scott Zeger
On behalf of the Department of Biostatistics

DEPARTMENT OF BIostatISTICS
SUMMARY OF RESPONSES TO QUALITATIVE SURVEY
2005

1. What are the open scientific questions that currently drive research in your field?

- Measurement: biomarkers; health of populations
- Causal inference/pathways
- Interactions: gene-environment; mixtures of exposures
- Rare adverse events – clinical data bases; population studies
- Infectious disease processes and models

2. What changes have occurred in the way the best studies in your field are currently being done and what has caused them?

- New technologies “-omics” creating lots of fishing expeditions
- More interdisciplinary work
- Large and more complex data sets
- Difficulty to recruit subjects
- Skepticism about instrumental variables
- Computational biology/modeling
- RCT reporting requirements

3. What new measurement techniques are having a major impact on how research is conducted in your field?

- Surrogate biomarkers for clinical trials
- Web surveys; audio CASI
- Toxicologic arrays
- Finer time resolution and particle composition
- Biotechnology measures

4. What new quantitative methods are increasingly popular in your field?

5. What quantitative methods does your group seek to develop further expertise in during the coming period?

- Marginal structural models, instrumental variables
- Latent variable models: hierarchical, longitudinal
- Multi-level; growth-curve models
- Agent-based computational models of epidemics
- Network structure analysis
- Time series models
- Validation of quantitative molecular biological measures
- Bayesian models that incorporate prior knowledge
- Percentile regression

6. How can the Department of Biostatistics be more supportive of your research program?

- Collaborate on substantive research
- Biostatistical challenges are at the core of some of the epidemiologic studies of the future. ... Is there a need for more generalists?
- More broadly advertise your working groups and open them to more faculty
- Say "Yes" more when asked to collaborate
- Collaboration on infectious disease modeling
- Already excellent (thanks Jim)
- Biostat faculty working on -omics problems should interact at a more global level with investigators who have a broader perspective... so time is not wasted... (with) poor quality data
- Faculty ... feel they don't get much for the effort (\$)
- Collaboration on medication error data bases
- Analysis of expenditure data
- Make two-term course into one-term course

8. What else would you like to tell us to make us a better department and/or more useful to your group?

- Biostatistics facility charge
- Publication bias
- Who does microarray analyses – we are going outside
- You do a great job (thanks Jim). Film Wall of Wonder presentations
- Your faculty attend our lab meetings
- Learn to provide highly specialized advice and also be a generalist
- Need more faculty
- Include statistical control theory in the curriculum
- Dept collaborate (with HPM) in a series of evaluation courses
- Make EBEG more accessible to non-statisticians

**DEPARTMENT OF BIostatISTICS
RESPONSES TO QUALITATIVE SURVEY
2005**

Colleagues who responded	
Alvaro Munoz	Epidemiology
Ann Klassen	Health Policy & Management
Brian Schwartz	Environmental Health Sciences
David Bishai	Population & Family Health Sciences
Donald Burke	International Health
James Tielsch	International Health
John Groopman	Environmental Health Sciences
Jonathan Samet	Epidemiology
Laura Caulfield	International Health
Laura Morlock	Health Policy & Management
Patrick Breyse	Environmental Health Sciences
Roger McMacken	Biochemistry & Molecular Biology
Terry Brown	Biochemistry & Molecular Biology
David Holtgrave	Behavior and Health
Colleagues who have not yet responded	
Alan Scott	Molecular Microbiology and Immunology
Chris Forrest	Health Policy & Management
Dani Fallin	Epidemiology
Diane Griffin	Molecular Microbiology and Immunology
Ellen MacKenzie	Health Policy & Management
Josef Coresh	Epidemiology
Michele L. Dreyfuss	Population & Family Health Sciences
Robert Blum	Population & Family Health Sciences
William Eaton	Mental Health
Cecile Pickart	Biochemistry & Molecular Biology
Michele Cooley	Mental Health

**DEPARTMENT OF BIostatISTICS
BREAKOUT GROUP LIST
2005**

Group 1	Group 2	Group 3	Group 4
Address Questions 1 and 2	Address Questions 3 and 4	Address Questions 1 and 2	Address Questions 3 and 4
Meet in the Tea House Room	Meet in the Rose Garden Room	Meet in the Cocoa Inn Room	Meet in the Garden Terrace West Room
Aristide Achy-Brou Mary Joy Argo Brian Caffo Gary Chan Howard Chang Frank Curriero Sandra Eckel Jay Herson Yen-Yi Ho Brendan Klick Dongmei Liu Yi-Chun Ouyang Ingo Ruczinski Rick Thompson Suyan Tian Lei Zhang	Ming-Wen An Karen Bandeen-Roche Ciprian Crainiceanu Chongzhi Di Francesca Dominici Sorina Eftim Mike Griswold Yi Huang Frank Hurley Sevasti Kohilas Fan Li Ani Manichaikul Kenny Shum Wenyi Wang Zhijin Wu Scott Zeger Hongling Zhou	Benilton Carvalho Leena Choi Snaebjorn Gunnsteinsson Jeffrey Hung Rafael Irizarry Rongheng Lin Xianghua Luo Tom Louis Debra Moffitt Georgiana Onicescu Luu Pham Fernando Pineda Stacey Rowuls Dan Scharfstein Chi Wang Yijie Zhou	Ron Brookmeyer Chao-Ling Chang Lijuan Deng Brian Egleston Jody Gatuso Hongfei Guo Elizabeth Johnson Yun Lu Jing Ning Stephanie Panichello Giovanni Parmigiani Rob Scharpf Shu-Chih Su Zhiqiang Tan Mei-Cheng Wang Xiaojun You

**DEPARTMENT OF BIostatISTICS
QUESTIONS TO ADDRESS
2005**

1. What are the most exciting emerging opportunities in public health and biomedical research for the next 3-5 years?
2. What are the associated statistical research topics?
3. What type of resources do we need to build expertise in emerging research topics?
4. What are the best opportunities to diversity funding for the department?

**DEPARTMENT OF BIostatISTICS
STUDENT AND FACULTY POSTER PRESENTATION SESSION
2005**

Leena Choi

Comparison of algorithms in PK/PD modelling.

Sorina Eftim

Spatial Confounding in Studies of Long Term Effects of Air Pollution.

Brian Egleston

A causal inference perspective on investigating mediation: Does sunlight exposure mediate the effect of eye-glasses on cataracts?

Hongfei Guo

Modelling differentiated treatment effects for multiple outcomes data.

Yi Huang

Average Treatment Effects (ATE) on Binary Outcomes: Measures, Collapsibility, Estimation by Propensity Scoring.

Elizabeth Johnson

Effects of Labor Interventions on the First Stage of Labor

Hormuzd Katki

Survival Analysis of Stratified Case-Cohort Studies to Estimate Relative, Absolute, and Attributable Risks, Using the R Software CaseCohort().

Brendan Klick

Avian species of the Niagara Frontier Region: Seventy years of changing abundances. Analysis of count data using additive models.

Fan Li

Are covariates covariate?-- A study of their role in linkage analysis using affected-sib-pairs.

Rongheng Lin

Ranking USRDS provider specific SMR with loss function based ranking method.

Yun Lu

Potential Application of Hidden Markov Model in the Quantification of Fetal Heart Rate and Fetal Movement Association.

Xianghua Luo

Recurrent event models in the presence of a failure event: comparison and inference.

**DEPARTMENT OF BIostatISTICS
STUDENT AND FACULTY POSTER PRESENTATION SESSION
2005**

Ani Manichaikul

Don't use the bootstrap for QTL mapping.

Jing Ning

Bivariate recurrent event process: modelling and inference.

Rob Scharpf

When should one subtract background fluorescence in two color microarrays?

Kenny Shum

Robust estimation of the mean of a positive random variable: an application to medical expenditure.

Wenyi Wang

Validation of Panpro - A Mendelian Prediction Model of Pancreatic Cancer Risk.

Yue Yin

An Infectious Disease Model for Maryland 1918 Influenza Data.

Xiaojun You

Statistical determination of the length of quarantine periods in an epidemic.

Yijie Zhou

Multi-level Models for Investigating Racial Disparity in Mortality and Socioeconomic Status in the Medicare Population.

**DEPARTMENT OF BIOSTATISTICS
BIOSTATISTICS RETREAT SCIENTIFIC PRESENTATIONS
2005**

2:00 – 2:45 PM: Section 1.

Daniel Scharfstein

Inferences and Decisions in the Presence of Non-Identifiability

Zhiqiang Tan

Estimation of Causal Effects Using Instrumental Variables

Scott Zeger

On Smoking-Attributable Death, Disease and Dollars in the U.S.

Francesca Dominici

Statistical Models for Large Spatio-Temporal Databases: Estimating Excess Number of Hospitalizations for Cardiovascular and Respiratory Diseases Attributable to Fine Particles and their Medicare Costs

2:45 – 3:00 PM: Break

3:00 – 3:45 PM: Section 2.

Giovanni Parmigiani

Integrative correlation: a tool for exploring cross-study reproducibility in high dimensional data

Mei-Cheng Wang

Analyzing recurrent longitudinal data

Ciprian Crainiceanu

Prediction versus Estimation in measurement error models

Karen Bandeen-Roche

The Use and Usefulness of Latent Variable Models

3:45 – 4:00 PM: Break

4:00 – 4:45 PM: Section 3.

Brian Caffo

fMRI and the Stroop Exam

T. A. Louis & Y. Yin

Bayesian Melding

Ingo Ruczinski

Stuff I am working on these days

Rafael A. Irizarry

Better Data are Better than Better Models