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 Dinning 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) Bandeen-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

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

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* Total Direct	1,902	1,637	2,103	2,203	1,758
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	*Modeling Biomedical Data and the Foundations of Bioequivalence*	Asst Prof Vanderbilt U Dept of Biostatistics	6
Michael Griswold	04-05	Zeger	*Complex Distributions, Hmmmm… Hierarchical Mixtures of Marginalized Multilevel Models*	President Griswold Consulting	7
Dongmei Liu	04-05	Parmigiani	*Application of Hierarchical Models in Microarray Data Analysis: Screening for Differentially Expressed Genes and Making Inference on Functional Classes*	Research Fellow London School of Hygiene & Tropical Medicine Dept of Infectious & Tropical Diseases	5
John Robinson	04-05	Zeger	*A Hierarchical Multivariate Two-Part Model for Profiling Providers' Effects on Healthcare Charges*	President John W. Robinson, MD, PhD, LLC	9
Michelle Shardell	04-05	Scharfstein	*The Analysis of Informatively Coarsened Discrete Time-to-Event Data*	Asst Prof U of MD Dept of Epi & Preventive Medicine	5
Ravi Varadhan	04-05	Frangakis	*The Role of the Design, Analysis, and Computation in Addressing Aetiology in Three Types of Studies in Public Health*	Asst Prof JHU Dept of Medicine	6
Zhijin Wu	04-05	Irizarry	*Probe Level Models for DNA Microarrays*	Asst Prof Brown U Ctr of Statistical Sciences	5
Weimin Chen	03-04	Broman	*Robust Quantitative Trait Linkage Analysis in Extended Human Pedigrees*	Postdoctoral Fellow U of Michigan Ctr for Statistical Genetics	4
Wesley Eddings	03-04	Rohde	*Topics in the Philosophy of Statistics: Methods, Data, and Theory*	Asst Prof Birmingham-Southern College Div of Sci & Math	6
Nikhil Gupte	03-04	Brookmeyer	*Statistical Models and Methods for Mother to Infant HIV Transmission Studies*	Data Manager/Statistician Johns Hopkins Department of Medicine, Division of Infectious Diseases	5

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.

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.

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 epidemioilogic 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 Breysse	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

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

Alvaro Munoz	• Parametric methods for time-to-event data • Competing risks as removals, not censored • Less data as fast progressor should not be penalized • Adverse events in chronic HIV infection

Ann Klassen	• Here is a current PA from NIH that presents a vehicle I would use for my spatial measurement research - and in such a grant, partnership with Frank or other spatial methodologists would be highly desirable. So I am forwarding it to you as an example of the kinds of research questions we would look to biostatisticians to be our partners on..... http://grants1.nih.gov/grants/guide/pa-files/PA-05-090.html
	  • Integration of behavioral, social, and biological data
	  • Measurement issues in research relating to diverse populations
	  • Measurement issues in studying potentially sensitive behaviors
	  • Measurement issues concerning ethics in research

Brian Schwartz	• This is a large question, but here is a brief answer. Understanding health effects of lower and lower doses; disentangling complex causal pathways; gene-environment interaction; development and validation of biomarkers of exposure, dose, susceptibility, and health effects; development and validation of new methods to assess health in populations (e.g., especially imaging techniques, 'omics, "latent" health constructs that cannot be directly measured); global environmental change; health effects of land use and urban sprawl; application of SNP, gene expression, and metabolomic technologies to population studies (large N and large P); definition and measurement issues (e.g., what is the built environment, what is the social environment, what is the best way to capture the entire pool of cognition-threatening toxicants); separating acute effects of recent dose from chronic effects of cumulative dose.
David Bishai	• Identifying causal treatment effects • Testing predictions from game theory using both simulation and experiments
Donald Burke	• How do epidemic infectious diseases arise • What are the strategies for, and longer term feasibility of, global eradication of diseases • How are human social structures organized, and can they be represented as canonical networks • Why is there person to person variability in susceptibility to infectious diseases

James Tielsch	• Many as I work in a variety of areas. The following are important
	  1. What are the genetic bases for the chronic ocular disorders that cause blindness and visual impairment? These include cataract, macular degeneration, glaucoma, etc. The genetic studies to date have been unimpressive in the attributable fraction of disease explained by genetic factors. How do these genes interact with environmental factors and which environmental exposures are important besides age.
	  2. How can we assess the micronutrient status of populations and indivduals? Important nutrients include vitamins A, E, C, and the minerals iron, zinc, selenium and iodine. Our current status indicators are not very good at the individual level and barely adequate at the population level. How can we estimate the impact of supplementation or food fortification when we have only limited information about the distribution of deficiency in the population? How can we account for this heterogeneity of status when trying to summarize information across studies (meta analysis)?

John Groopman	• There are two major research questions that we are addressing;
	  1. Role of gene-environment, primarily aflatoxin and p53, in human liver cancer and the role of gene-environment-vector (p53, aflatoxin, HBV/HCV infection) in human liver cancer.
	  2. Validation of intermediate biomarkers in chemoprevention trials in high-risk populations.

Jonathan Samet	• risks of low levels of air pollution • determining features of mixtures that are associated with toxicity • trans-disciplinary work on mechanisms

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

Laura Morlock	• In the field of patient safety an important activity is adverse event reporting. Some adverse events involve patient harm, while the great majority of reports in most reporting systems involve the perception on the part of the reporter of elevated risk, but no harm (often called "near misses"). Adverse event reporting in health care is modeled after the systems in aviation and other high risk industries where near miss reporting is credited with substantial improvements in safety. This is partly because adverse incidents involving harm are—fortunately—rare events. The usefulness of near miss reporting in health care is currently an act of faith, while consuming substantial resources for reporting, analysis and feedback. Open questions include whether patterns (e.g., contributing factors) are similar for adverse events with and without harm, and the extent to which the patterns identified in near miss reports provide information useful for patient safety interventions. • We are examining these issues in intensive care units around the country which are participating in our web-based ICU adverse event reporting system (under the leadership of Peter Pronovost, MD, PhD). • We have just signed an MOU with the U.S. Pharmacopeia to help them examine these and other issues. They are responsible for the MedMARX reporting system for adverse events involving medications. Over the past five years approximately one million "medication errors" have been reported to this system.
Patrick Breysse	• the questions are exposure related. much of my research today is related to measures of air pollution and their relationships to health (eg asthma) and to toxicological mechanisms.
Roger McMacken	• This query is too broad to be answered for the general fields of biochemistry and molecular biology, so I will reply for the more specialized fields related to my own research: a) initiation and regulation of DNA replication; and b) mechanisms of action of molecular chaperones in remodeling of macromolecular assemblies. • Important open questions for studies of the initiation of DNA replication: 1. How do chromosomal replication initiators promote opening of the DNA duplex at replication origins? 2. What molecular mechanisms are involved in loading of replicative DNA helicases at replication origins? 3. By what mechanisms do viruses recruit the replication machinery of host cells to replicate viral chromosomes? 4. What mechanisms insure that each segment of a chromosome is replicated once and only once per cell division? • Important open questions for studies on molecular chaperone action: 1. How do molecular chaperones distinguish between native and unfolded protein substrates? 2. How do Hsp70 and Hsp40 chaperones cooperate to aid folding of unfolded proteins or transport of unfolded proteins across cellular membranes? 3. What molecular mechanisms are involved in the remodeling of complex, multiprotein substrates by molecular chaperones?
David Holtgrave	• A very urgent need is to utilize statistical techniques appropriate for analyzing outcome data from trials in which multi-level interventions are being assessed. For instance, one might intervene simultaneously at the individual, dyadic and community level in an HIV prevention trial - analyzing the data from such a study would pose clear challenges. The analysis of data from multi-level intervention trials may be the most difficult challenge we currently face.

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

Alvaro Munoz	• Effective therapies in HIV infection have changed the epidemiology of HIV from a lethal infectious disease to a chronic condition • Cohorts of HIV are now cohorts of CVD, chronic hepatitis, metabolic complication
Brian Schwartz	• Application of new technologies; more multi-disciplinary and inter-disciplinary work, working at the boundaries of existing disciplines; need for larger and larger sample sizes, more complex sets of dependent and independent variables (complex causal pathways) that are inter-correlated. Driven by availability of technology, new thinking about disease causation, funding agencies. One ongoing challenge is the increasing difficulty to identify, select, and recruit study subjects.
David Bishai	• Higher levels of skepticism for identification when using instruments. Preference for natural experiments to achieve identification in social research.
Donald Burke	• Computational modeling of microbial emergence • Computational modeling of vaccine trial design • Molecular epidemiology of microbes
James Tielsch	• Little has changed in the past 10 years. In ophthalmic epidemiology, there is growing frustration with being unable to identify factors that can explain an important fraction of the disease in a population (besides age).
John Groopman	• The development of –omics methods using instruments ranging from mass spectrometry to micro and tissue arrays has vastly increased the volume of data that now needs to be analyzed. Better study designs are needed to get us out of the “fishing expedition” conundrum.
Jonathan Samet	• increasing size of studies • more sophisticated outcomes • incorporation of new technologies
Laura Caulfield	• RCT reporting requirements
Laura Morlock	• Adverse event reports in health care currently are “counts.” While useful, these also need to be converted to rates. How to construct denominators that most appropriately measure risk exposure is another open question.
Patrick Breysse	• new measurement techniques have created the opportunity to assess exposure with very fine time resolution. the question becomes what is the appropriate time averaging to relate to different health outcomes and is a simple time-weighted average the best metric for predicting adverse outcomes • the growth of "omic" tools has changed how we look at exposure relationships and how we investigate the impact of the environment • collaborating with toxicologists to create more "environmentaly relavent" studies using real-world exposure models that are more closely linked to epidemiologic studies.
Roger McMacken	• Today and historically, the best work in these fields is carried out by the top rank biochemists, enzymologists, and molecular biologists. However, studies in these fields have been facilitated by the development of new technologies that enable analysis of single molecules or real time analysis (by plasmon resonance approaches, e.g., Biacore) of complex protein-protein or protein-DNA interactions. Additionally, chromatin immunoprecipitation (ChIP) assays have proven very instructive in analysis of the DNA-binding events that occur at eukaryotic DNA replication origins prior and during the initiation of chromosomal DNA replication.
David Holtgrave	• CDC has recently promulgated methodological standards for the reporting of quasi-experimental studies in public health (this is the CDC TREND group). It seems to me that some researchers doing such quasi-experimental work and who would like to meet CDC's methodological guidelines could use some statistical support in this arena. • As we come to recognize that public health interventions occur at multiple levels, more and more attention is being placed on societal level (including policy and legal) interventions. In some cases, assessing the impact of naturally occurring changes in policies and laws would require the use of interrupted time series analysis; this technique seems rather underutilized in the behavioral and social sciences and could use further support to broaden its appropriate use.

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

Alvaro Munoz	• A large number of markers which hope to operate as surrogate of clinical outcomes
Brian Schwartz	• Imaging, 'omics technologies, remote sensing, GIS, new personal sampling devices to assess behavior and the environment in real time; some new hindrances to study subject identification, selection, and recruitment.
David Bishai	• Audio CASI • Web Surveys
Donald Burke	• 'omics and immune responses: genomics, transcriptomics, and proteomics, as they affect immunity • Measures of genomic similarity between microbes
James Tielsch	• No major ones that I can think of except for genome wide scans etc.
John Groopman	• As described above –omics (genomics, proteomics, metabolomics and phenomics) are driving the field. Each of these technologies remain un-validated and cross-talk across these platforms are more myth than reality.
Jonathan Samet	• Air pollution measurements of increasing specificity • More sophisticated toxicologic arrays
Patrick Breysse	• the ability to collect time resolved Paticulate Matter data
Roger McMacken	• As discussed under item 2, newly developed technologies permit more precise and more sensitive measurements of complex biological reactions and structures. These include biosensor (Biacore) analysis, stopped-flow fluorescence analysis of reaction rates, isothermal titration calorimetric (ITC) analysis, and improved methodologies for structural analysis of large macromolecular complexes by crystallography or nuclear magnetic resonance.
Terry Brown	• Gene expression analyses, microarrays and bioinformatics • Quantitative real-time PCR
David Holtgrave	• Some measurement issues that we've faced recently at Emory include the measurement of very sensitive information (especially sexual and drug use behaviors), the self-reporting of biomedical conditions (e.g., the self-reporting of HIV serostatus when the study can not afford to measure HIV directly), and the measurement of behaviors in difficult street settings (where the setting could impact the attention and mood of both researcher and respondent). • We also struggle with the issue of measuring social constructs at the state level. A major issue at the moment is how to best measure social capital at both the community and state level. As we broaden our focus from individual behavior to societal determinants and interventions, understanding how to measure such constructs at increasingly broad areas is an important challenge.

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

Alvaro Munoz	• Marginal Structural models • Instrumental variables
Brian Schwartz	• For measurement, those above. For analysis – all sorts of multivariable and multivariate methods; latent variable models (factor, path, structural); hierarchical models; longitudinal models.
David Bishai	• SEMs • Multi-level models • Growth Curves
Donald Burke	• Agent-based computational modeling of epidemics • Network structure analyses • Epidemic time series analyses
James Tielsch	• None come to mind as hot or popular.
John Groopman	• In the –omics world much of the analyses remain qualitative and not quantitative. Internal standards are need to make these tools quantitative and the validation of these quantitative techniques have yet to be done in biological models or in human investigations.
Jonathan Samet	• Larger air pollution data sets analyzed with approaches to characterize ? • Time-series models • Multi-level approaches
Patrick Breysse	• basian methods for retrospective exposure assessments that incorporate expert judgements.
Roger McMacken	• See response to question 3. Measurement and quantitation are synonymous with respect to new, cutting-edge instrumentation for study of biological molecules and systems.

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

Alvaro Munoz	• See #1

Brian Schwartz	• All those above. Also working with Tom Glass on new personal sampling devices to measure social interaction, cognitive function, physical activity, and location in space over time. Want to extend this to work to develop such personal sampling devices to measure the built environment.
David Bishai	• See #4
Donald Burke	• Computational modeling of epidemics • Computational modeling of vaccine trials • Analysis of transcriptome changes in peripheral blood cells in infectious diseases and post-vaccination, to devise predictors of favorable and unfavorable outcomes.
James Tielsch	• Time-dependent covariate adjustment and interpretation.
John Groopman	• We are primarily an analytical chemistry based program. These tools have served us well, but the throughput will always remain low. Therefore, we need to have better study design to maximize our investigations within the confines of highly sensitive but low throughput methods.
Jonathan Samet	• Further evaluation of spatial-temporal modeling
Laura Caulfield	• Longitudinal methods • MSM • Percentile regression, intermediate outcomes • Copying with errors in variables
Patrick Breysse	• see #4
Roger McMacken	• We will focus our quantitative approaches in crystallographic, ITC, stopped-flow fluorescence, and biosensor analyses of the replication reactions and molecular chaperone-dependent events under investigation in our laboratory.
David Holtgrave	• There are a number of methodological tools available in the field of marketing that could be exploited in the behavioral and social science arenas of public health, however, they are seemingly not widely known or used in public health. These tools include conjoint analysis, information tracing, and response tracking (in mass media message evaluation). Further utilization of such methods could be very useful in public health (a student of mine just did a thesis using information tracing techniques and wound up being nominated for a science prize largely on the basis of the uniqueness of his application of a rather basic marketing research strategy in public health).

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

Alvaro Munoz	• Collaborate on substantive research
Brian Schwartz	• I have a need for more support. Access to and production by biostatistics faculty have been increasingly difficult as my needs have grown and their collaborations have increased. I have several long-term collaborations with the department that are highly valued, and some new ones that are developing, but have a need for more (thus our desire to recruit into EHS). Biostatistical challenges are at the core of some of the epidemiologic studies of the future. When a study requires someone who knows about longitudinal, multi-level, latent variable, spatial statistics and complex neuroimaging methods all at once, there are challenges to efficient progress. It seems biostatistical knowledge, like much scientific knowledge, exists in silos and thus bringing expertise to bear is very difficult logistically, financially, and scientifically. Is there a need for more generalists? Is there a way to train faculty to have broader skill sets? Is there a way to teach methods that can be broadly applied across a range of similar statistical problems (whether the data complexity is over time, over levels, over space, or over all three)? It would also be valuable to more broadly publicize your working groups and open to other faculty.
David Bishai	• Develop capacity to say "Yes" more when asked to collaborate
Donald Burke	• Collaborations on analytic modeling of infectious disease epidemiology • Collaborations on agent based simulations • Collaborations on analyses of transcriptome changes in peripheral blood cells in infectious diseases and post-vaccination
James Tielsch	• Already receive excellent support and collaboration.
John Groopman	• It would probably benefit faculty interested in –omics research to interact at a more global level with investigators that have a broader perspective on the goals of epidemiologic and experimental studies so time is not wasted developing advanced analysis methods for poor quality experimental data.
Jonathan Samet	• doing ok
Laura Caulfield	• Faculty need statisticians on their grants but I think many don't feel they get much for the effort ($)
Laura Morlock	1. We are hoping that there will be some interest in collaborating with us in exploring the medication error data base described above. 2. On other projects we frequently work with highly skewed expenditure data or health services utilization data. A typical pattern might be for 35% of the cases to be zero, and 5% of the cases to be in a long right tail of the distribution which might account for 50% of the total expenditures/utilization. Typically we use two stage models; in the second stage the distribution is usually normalized in some way. I was very interested in Kenny Shum's discussion (preliminary oral exam) regarding the possibilities of three stage models for handling these types of data—I really think this could be an important new measurement technique!
Patrick Breysse	• from my experience the dept has been very supportive. I look forward to having F. Curriero in the dept and hope biostats will be supportive of his development
Roger McMacken	• As far as I am concerned, the Biostatistics Department is already doing quite well in offering course work that is relevant to the needs of my students. However, in general from the BMB perspective, I feel that the current two-term course designed for lab science students could be shortened to one term (for most students) and focused more directly on issues most relevant to measurements in the laboratory setting (Poisson distributions, Gaussian distributions, errors in measurement of biological reactions, curve-fitting methods, statistics of small sample sizes [for those students that work with animals], etc. (I realize that not all of these mentioned problems involve classical statistical approaches) More complex statistics relevant to analysis and interpretation of DNA or protein microarray data, or analysis of population-based data (which is of less interest to most BMB students), might best be offered in a second term. A course set up along these lines would likely attract a higher proportion of BMB PhD students.

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

David Holtgrave	• Statistical support is always needed in psychometrics. I am always shocked by how many papers I review purport to measure behavior but pay no attention to the reliability and validity of the measures. It seems that authors are aware of the need to attend to reliability and validity but do not know how to do so in a quantitative and rigorous way. • Additionally, the techniques of factor analysis, cluster analysis and multi-dimensional scaling are highly relevant to social sciences applications in public health but are also vastly underutilized. • I have a very selfish request: In the area of cost-effectiveness analysis, there are new Bayesian methods being developed and although I would be interested in using them, they are a bit beyond my statistical limits. May I please get in line for some Bayesian consultations in cost-effectiveness analysis?

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

Alvaro Munoz	• Cole SR, Li R, Anastos K, Detels R, Young M, Chmiel JS, Munoz A. Accounting for leadtime

	in cohort studies: evaluating when to initiate HIV therapies. Stat Med. 2004 Nov 15;23(21):3351-63. • Detels R, Munoz A, McFarlane G, Kingsley LA, Margolick JB, Giorgi J, Schrager LK, Phair JP. Effectiveness of potent antiretroviral therapy on time to AIDS and death in men with known HIV infection duration. Multicenter AIDS Cohort Study Investigators. JAMA. 1998 Nov 4;280(17):1497-503. • Li X, Buechner JM, Tarwater PM, Muñoz A. A diamond-shaped equiponderant graphical display of the effects of two categorical predictors on continuous outcomes. The American Statistician 2003;57:193-99
Brian Schwartz	• Your faculty already knows about the key papers. I can provide a list if requested.
David Bishai	• Rosenzweig, M, and Wolpin K. 2000. Natural 'Natural Experiments' in Economics. Journal of Economic Literature 2000; 38(4), 827-874
Donald Burke	• Ferguson NM, Fraser C, Donnelly CA, Ghani AC, Anderson RM. Public health. Public health risk from the avian H5N1 influenza epidemic. Science. 2004 May 14; 304 (5673): 968-9.
John Groopman	• Kensler T, Qian G-S, Chen, Groopman J. Translational Strategies for cancer Prevention in Liver, Nature Reviews Cancer. May 2003, 3(5) 321-329. • Nicholson J. Connelly J. Lindon J. Holmes E. Metabonomics: a platform for studying drug toxicity and gene function. Nature Reviews Drug Discovery. February 2002, 1(2) 153-161.
Laura Morlock	• One useful (but a little dated) discussion is R.J. Lilford, et al., "The Measurement of Active Errors: Methodological Issues" in the Journal of Quality and Safety in Health Care 2003;12(suppl II):ii8-ii12. (in the patient safety field "active errors" refer to something that has occurred, while "latent errors" are essentially accidents ready to happen once a triggering event occurs)
Patrick Breysse	• Ramachandran G. Retrospective exposure assessment using Bayesian methods. Ann Occup Hyg. 2001 Nov;45(8):651-67.
Roger McMacken	• I can't think of a paper that encapsulates and incisively extols the most important trends that I have outlined above.

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

Brian Schwartz	• The Biostatistics facility charge seems to be a bit difficult to justify if my projects are not using your facilities. It is a tax that puts your department's collaborators in a very awkward position.

David Bishai	• If rationale is not there, there will likely be "publication bias", although natural instincts of good scientists would be to correct this bias and accept the consequences.
Donald Burke	• We still haven't solved the problem of who does the microarray analyses. It's a bit difficult for most investigators to pick up on their own, and there's a steep learning curve for students, yet biostats doesn't want to be in the business of actually doing the analyses in collaboration with the field/bench scientists. This forces us to go outside the SPH.
James Tielsch	• I think you guys do a great job. Only wish I did not miss the Wall of Wonder presentations. Any chance you can record them in addition to the PP slides?
John Groopman	• It would be good to have some of the faculty who are interested in our research to attend our lab meetings.
Jonathan Samet	• Think about how to address highly specialized areas (e.g. air pollution) while still providing general support. What niche areas can be addressed? How many? How are selections made?
Laura Caulfield	• I think you have a great department. Need more faculty as it is often hard to find collaborators on grants. Often hard to figure out who to ask to collaborate with.
Laura Morlock	1. We are also very involved in performance improvement activities in health care organizations. Statistical process control can be an important technique for data that are often collected quarterly. I realize this approach was developed in the field of engineering—but I'm wondering if it should be included in our curriculum? (perhaps this is more related to the second point below). 2. I think there are many students who define themselves as non-researchers and who resist taking the 620 series. Many of these students will never participate in research projects, but will be involved all their professional lives in evaluating various projects and programs. Many of the statistical methods are similar, but are often used within a different conceptual framework. We are starting to think through a sequence of evaluation courses for master's and DrPH students. This needs to be an interdepartmental effort, and we are hoping that Biostatistics will be interested in collaborating.
Patrick Breysse	• make the EEBG working group more accessible (in terms of content) to non-biostatiticains
Roger McMacken	• The Biostat Dept. is already doing a superb job in education and research. However, the Biostatistics Department might consider some partial revision of the Biostat courses for lab science students along the lines suggested above … but, of course, only if other laboratory investigators agree with this course of action.
David Holtgrave	• I would also be interested in your thoughts about introducing students to randomization tests (ala Edgington). I found them very useful in some methods work I did on developing statistical inference tests in multi-attribute utility analysis. At least at Emory, we never teach students about this widely applicable and useful technique. Are we doing them a disservice but not exposing them to what is a rather fundamental and flexible technique? • Summary: In some cases, I think we need ever more sophisticated techniques (such as analyzing data from multi-level interventions). However, in some cases, I feel that there are many useful tools out there (such as multi-dimensional scaling) that simply are not taught or discussed much in public health schools; this is really not an issues of developing new techniques but trying to truly exploit a whole array of extant quantitative psychology, sociology and marketing research methods.

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 Sandrah 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 Stacee 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
Comparision 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 funtion based ranking method.

Yun Lu
Potential Application of Hidden Markov Model in the Quantification of Fetal Heart Rate and Fetal Movement Asscociation.

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