THE FUTURE OF STATISTICS AND BIOSTATISTICS

Thomas A. Louis PhD, Department of Biostatistics
Johns Hopkins Bloomberg School of Public Health
www.biostat.jhsph.edu/~tlouis/
tlouis@jhsph.edu

"Prediction is a very delicate art, especially with regard to the future"
Science $\approx$ 1980

MY PREDICTIONS FROM 1999 & 2002

"SPOT ON"

- The future lies ahead
- These are exciting and challenging times
- New opportunities and challenges will emerge
- The only certainty is uncertainty

ESSENTIALLY CLUELESS

- We'll move away from PCs to workstations

VIEWS OF OTHERS

Statistical Challenges and Opportunities for the Twenty-first Century

B Lindsay, J Kettenring, D Siegmund, eds., August 2004, NSF

"Statistics is the science of learning from data"

We need to:

- Promote understanding of statistical science
- Develop more flexible funding models
- Strengthen the core of statistics research
- Improve support of multidisciplinary research activities
- Develop new models for statistical education
- Accelerate recruitment of the next generation
- Increase support for and autonomy of the NSF statistics program

A Cox Model for Biostatistics of the Future, Zeger SL, Diggle PJ, Liang K-Y
(2004 or 2005, in the volume honoring D. R. Cox on his 80^{th} birthday)

- The 20^{th} century has produced lists of genes and proteins; the next priority is to determine their functions
- Having quantified tens of thousands of gene or protein expression levels, the question is which values best predict the incidence or progression of disease or the efficacy of a novel treatment
- Similar problems are common in image analysis
- The increasing amount of data and computing power will make possible a new level of public health surveillance

C.R. RAO, AmSTAT News, September 2004

- Statistical Methodology is mostly model-based
- Too much emphasis is laid on tests of significance
- There is no standard way of expressing uncertainty
- Most departments don't offer courses in sample surveys or design of experiments
- Students don't have knowledge of other disciplines
- The role of the statistician should not be merely that of a technician
- Statistics does not develop from statistics
- The future of statistics lies in communication of statisticians with other researchers

TOPICS

- Methods Futures
- Education: Undergraduate, graduate, continuing
- Mission and Organization
- Sociology & Politics

SOME METHODS FUTURES

- Statistical modeling with "P, n"
- Gene expression including microarrays and SAGE
- Imaging
- Generalizations of PCA
  (Donoho at the 2004 JSM)
- Integrating remote sensed and and "ground truth" information
- Analysis of spatio-temporal information
- Computing with $N \times P$

MORE METHODS FUTURES

- Hierarchical Models
  - (Hyper-prior) sensitivity
  - Goodness of fit
  - Influence of the model and the data on conclusions
- Causal Analysis
  - Multi-level flows and influences
  - Skepticism of reifying a hierarchical model
- Controlling/accommodating multiplicity
  - Gene-environment interactions where there are thousands of candidate relations

AND STILL MORE METHODS FUTURES

- Knowledge Discovery; Neural Nets
- CART
- Smooth, semi-parametric
- Wavelets
- Spatial analysis
- Geographic Information Systems
- Confidentiality, data disclosure
- Syndromic surveillance
- Embedding microsimulation (Agent-Based models) in a statistical framework
- Bayes, Bayes, ...

THE "UNION OF ALL METHODS"

- Complex, hierarchical, latent variable models, gene, environment, biomarkers of exposure, biomarkers of effect, health outcomes, brain scans/voxel analysis, cross-sectional and longitudinal, Berkson & Standard & Hybrid measurement error, parametric/non-parametric approaches, analysis of "objects," lots of exploration, unsupervised and supervised in complex ways, large datasets, ...

QUANTIFYING & COMMUNICATING UNCERTAINTY

Uncertainty will continue to play a key role in communicating risks and making decisions

- Heterogeneity is our friend, our job security, but it can also be our foe
- People want certainty; we deliver uncertainty
- Balancing acts:
  - Communicating the "best bet" while responsibly communicating the true uncertainty
  - Influencing behavior or opinions while protecting credibility of the process
- Need to educate stakeholders to insist on comprehensive uncertainty assessments

COMPUTING

- Astounding advances in speed, storage, cost, ease of use
- Ability to implement:
  - The bootstrap
  - MCMC
  - Bayesian analyses
  - Genetic linkage
  - Neural Nets
  - Image reconstruction
  - Simulations
  - Wavelets
  - ACE, MARS, GAMS, ...
  - Designs for complicated models

All of the foregoing should rely on statistics and statisticians

EDUCATING STATISTICIANS

Our future depends on educating the next cohort of masters and doctoral statisticians who:

- Understand statistical concepts & technology
- Engage in modern computing
- Appreciate Statistical Science and the role of statistics and statisticians
- Respect the broad spectrum of our discipline
  - from no-nonsense theory to no-nonsense applications
  - in industrial, governmental and academic settings
- Refresh knowledge via continuing education

RECRUITING & FUNDING STUDENTS

- Re-emergence of biostatistical training grants
- Opportunities to fund students on projects
- Need to attract additional students from the U.S.
- Need to rectify visa policy for foreign students

Educating collaborators and other professionals

- Big dose of biostat. for MPH and others
- Development of intermediate-level courses
- Dealing with teaching of statistics by other departments
- Distance education and short courses
- Success reaps many rewards, creating collaborators and political support
- Failure can lead to atrophy or demise

SCIENTIFIC COLLABORATION

Interdisciplinary research is very important, but it is difficult to conduct without disciplines!

- We must generate discipline-based researchers and educators, many of whome will collaborate
- We must imbue a respect for collaborations and expose students to collaborations
- We must maintain a disciplinary focus

ORGANIZATIONAL CHALLENGES & OPPORTUNITIES

- Some departments have closed
- Money used to be quite plentiful
  - Now, it's not quite so plentiful
- Indirect cost recovery funds were very flexible
  - Now, they are a net loss
- Enlightened tuition sharing produces flexible funds from all forms of education

"Since you're teaching a lot,
I can pay for several professional meetings"

ORGANIZATIONAL CHALLENGES & OPPORTUNITIES
CONSULTING

- Consulting is one of the chronic organizational and mission challenges in biostatistics
- It is necessary, desirable and sometimes vexing
- Need to decide:
  - Who gets involved
    - Faculty at various levels
    - Students
    - Center-specific staff
  - How they get involved
  - Charges and core Financing

$\Longleftrightarrow$ Mission

Location, Location, Location (and scale)

- Proper organizational context for biostatistics:
  - SPH ideal, but other contexts can work
- Department size and activities
  - Clinical trials and field trials centers
  - Micro, small and big science

MISSION

- Need to have mix of statistics-specific and collaborative educational and research missions
- Need to view ourselves as and behave as equals
  - In industry, government and academe, statisticians who do not take leadership do not produce the best science or policy
  - In academe, such behavior will not produce the best statistical science or scientists

SOCIOLOGY

- In a recent *AMSTAT News*, Steve Stigler noted that statistics has a bright future, but that this prognosis does not necessarily apply to the ASA

I've morphed that to:

- Development and application of statistical and biostatistical methods will continue to burgeon, but it is not automatically the case that either will be dominated by statisticians and biostatisticians
- The strength of our discipline, that statistical concepts and methods are fundamental to much of science and policy, can also be a weakness

BIOSTATISTICS IS A BASIC SCIENCE

- Of Public Health and many other fields
- Needs increase; awareness increases more slowly

WE NEED TO BE ADVOCATES

- For biostatistical science and scientists
- To influence, but not monopolize
- To educate biostatisticians, collaborators and consumers in the broad concepts of our field
- To promote our competitive advantages
- To promote broad ownership of statistical issues
- To avoid the analogue of:

  "Only historians are allowed to reminisce"

WE NEED TO:

- Insist on proper roles and "rules of engagement"
- Communicate our professional roles and goals
- Be scientifically and socially responsible, but not shy
- Have high standards, while making sure to promote and support (bio)statistics and (bio)statisticians

STATISTICIAN ROLES

(page 148 in Marks)

A. Bradford Hill (1966)

"Once the clinician has grasped the simple techniques that have been brought to his aid, the statistician has no further part to play. Along with the old soldier he can fade away, contentedly if, sometimes, wistfully."

L. Lasagna (1955)

"It is currently fashionable in some circles to consider the clinician member of the team as some sort of minor excrescence, 'a fifth cousin about to be removed.' "

INVESTIGATIVE TEAMS
(Level 0 design)

- NINDS constituted a panel of statisticians to review "statistical issues" in two high-visibility stroke trials
- But, of course, nothing is purely statistical (though almost everything has statistical aspects!)
- Fortunately, we insisted on including clinicians
- A true team effort produced:

  Ingall T, O'Fallon M, Asplund K, Goldfrank L, Hertzberg V, Louis TA, Christianson TJH (2004). Findings from the reanalysis of the NINDS tissue plasminogen activator for acute ischemic stroke treatment trial. Stroke, 35: 2418-2424

  Full report at:

  <http://www.ninds.nih.gov/t-PA_review_committee>

ENCOURAGING TRENDS

- The R-enterprise
- Bioconductor
- BUGS and JAGS

DISCOURAGING TRENDS

- Patenting statistical procedures
- Secrecy

THE EFFECT OF 9-11

- All fields must deal with the need to be mindful of national security while engaging in free inquiry
- This poses an important challenge

see Donald Kennedy, Science, 305:1077

TRADITIONAL VALUES STILL APPLY

- Top quality information
- Top quality statistical designs & analyses
  - that address the science/policy issues
  - that respect data limitations and avoid space-age analyses of stone-age data
- Full assessment of operating characteristics of designs and analyses
- Appropriate reporting of results
  - In science/policy contexts
  - Communicating & interpreting uncertainty
- Central biostatistical roles in conceptual, technical, administrative, policy and political domains

SUMMARY

- The importance and impact of biostatistics in science and policy burgeons
- Challenges in applications energize methodologic development, enhancing applied activity, ...
- Exciting new opportunities complement those in historically active domains
- Computing advances empower complicated statistical analyses by biostatisticians and others
- Though public awareness of the vital role played by biostatistics and biostatisticians increases, it lags substantially behind reality
- Therefore, sociological and political factors are important determinants of our future

CODA

- With history created by people like Donna Brogan
- With the present in the hands of Emory Biostatistics faculty
- With the future in the hands of current and future Emory Biostatistics students

THE FUTURE OF STATISTICS AND BIOSTATISTICS IS SHINING BRIGHT