ENAR 2016 Presidential Invited Speaker

Xihong Lin, PhD

ENAR President Invited Talk

Chair and Henry Pickering Walcott Professor
Department of Biostatistics | Harvard University

Biostatistics, Biomedical Informatics, and Health Data Science: Research and Training

Biostatistics has played a pivotal role in both the development and success of basic science, public health, and medical research by developing statistical methods for study design and data analysis. Massive 'ome data, including genome, exposome, and phenome data, are becoming available at an increasing rate with no apparent end in sight. Examples include Whole Genome Sequencing data, large-scale remote-sensing satellite air pollution data, digital phenotyping data, and Electronic Medical Records. The emerging field of Health Data Science (HDS) presents biostatisticians with many research and training opportunities and challenges. It has propelled us to rethink our identity and niche and how we can properly position ourselves as a leader in HDS, especially in promoting and advancing statistical inference in health data science research and training. Success will both for biostatistics and for much of health and biomedical science that we effectively position ourselves together with bio- and medical informaticians, as leading health data scientists. There are countless of examples where the volume of available data requires new, scalable statistical methods and demand an investment in statistical research. These include signal detection, network analysis, integrated analysis of different types and sources of data, and incorporation of domain knowledge in health data science method development. Especially critical is training the next generation of health data scientists, which include not only providing broader training of health and biomedical researchers in sound statistical inference, but also that integrate computer and information science and machine learning into established biostatistical curriculum. Such enhanced training could include both didactic and EdX courses, but will require a careful balance of depth and breadth across areas. In this talk, I discuss some of the challenges and opportunities, and illustrate them using statistical genetics and genomics as examples.

Biography

Xihong Lin is Chair and Henry Pickering Walcott Professor of Department of Biostatistics and Coordinating Director of the Program of Quantitative Genomics at the Harvard T. H. Chan School of Public Health. She received BS in Applied Mathematics from Tsinghua University, China and PhD in Biostatistics from University of Washington. Dr. Lin's research interests lie in development and application of statistical and computational methods for analysis of massive genetic and genomic, epidemiological, environmental, and medical data. She currently works on whole genome sequencing association studies, genes and environment, analysis of integrated data, and statistical methods for massive health science data.

Dr. Lin received the 2002 Mortimer Spiegelman Award from the American Public Health Association and the 2006 COPSS Presidents' Award. She is an elected fellow of ASA, IMS, and ISI. Dr. Lin received the MERIT Award (R37) (2007-2015), and recently the Outstanding Investigator Award (OIA) (R35) (2015-2022) from the National Cancer Institute, which provides "long-term research support to experienced investigators with outstanding records of cancer research productivity who propose to conduct exceptional research." She is the contacting PI of the Program Project (P01) on Statistical Informatics in Cancer Research, and the T32 training grant on interdisciplinary training in statistical genetics and computational biology.

Dr. Lin was the former Chair of the COPSS (2010-2012) and a former member of the Committee of Applied and Theoretical Statistics (CATS) of the National Academy of Science. She is the Chair of the new ASA Section of Statistical Genetics and Genomics. She was the former Coordinating Editor of *Biometrics* and the founding co-editor of *Statistics in Biosciences*, and is currently the Associate Editor of *Journal of the American Statistical Association* and *American Journal of Human Genetics*. She has served on a large number of statistical society committees, and NIH and NSF review panels.

ENAR 2016 IMS Medallion Lecture

Peter J. Diggle, PhD

CHICAS, Medical School, Lancaster University

Model-Based Geostatistics for Prevalence Mapping in Low-Resource Settings

In low-resource settings, disease registries do not exist, and prevalence mapping relies on data collected through a finite, often spatially sparse, set of surveys of communities within the region of interest, possibly supplemented by remotely sensed images that can act as proxies for environmental risk factors. A standard geostatistical model for data of this kind is a generalized linear mixed model,

$$Y_i \sim \text{Bin}\{m_i, P(x_i)\}$$
$$\log[P(x_i) / \{1-P(x_i)\}] = z(x_i)'\beta + S(x_i),$$

where Y_i is the number of positives in a sample of m_i individuals at location x_i, $z(x)$ is a vector of spatially referenced explanatory variables and $S(x)$ is a Gaussian process.

In this talk, I will first review statistical methods and software associated with this standard model, then consider several methodological extensions whose development has been motivated by the requirements of specific applications.

I will focus in particular on prevalence mapping projects that have arisen in connection with pan-African control programs for onchocerciasis (river blindness) and lymphatic filariasis (elephantiasis). These vector-borne diseases are major public health problem in the wet tropical regions of the world, including most of sub-Saharan Africa. Multi-national control programs using mass administration of a protective drug, Mecitzan, have been very successful, with more than 60 million treatments to date over 19 countries. However, the programs has been hampered by the recognition that people heavily infected with a third disease, *Loa loa* (eyeworm) parasite, are at risk of severe, occasionally fatal, adverse reaction to Mectizan. Before the drug is administered in a community, it is relatively easy to estimate the prevalence of eyeworm infection, harder (and more expensive) under field conditions to estimate how many people are "heavily infected," one definition of which is that they as carrying more than 8,000 parasites per ml of blood. To address this problem we develop a joint model for community-level prevalence and the proportion of highly infected individuals in the community.

Biography

Peter Diggle is Distinguished University Professor of Statistics in the Faculty of Health and Medicine, Lancaster University. He also holds Adjunct positions at Johns Hopkins, Yale and Columbia Universities, and is president of the Royal Statistical Society (2014-2016)

Between 1974 and 1983 Prof Diggle was a Lecturer, then Reader in Statistics at the University of Newcastle upon Tyne. Between 1984 and 1988 he was Senior, then Principal, then Chief Research Scientist and Chief of the Division of Mathematics and Statistics at CSIRO, Australia. He has worked at Lancaster University since 1988, and held a joint appointment with the University of Liverpool from 2012 to 2015. Between 2004 and 2008 he held a UK Engineering and Physical Sciences Senior Fellowship.

Prof Diggle's research involves the development of statistical methods for spatial and longitudinal data analysis, motivated by applications in the biomedical, health and environmental sciences. He has published 10 books and more than 200 articles on these topics in the open literature.

He was awarded the Royal Statistical Society's Guy Medal in Silver in 1997 and is a former editor of the Society's Journal, Series B. In 2001 he was elected as a Fellow of the American Statistical Association. He was founding co-editor of the journal "Biostatistics" between 1999 and 2009, and is a Trustee for *Biometrika*. He has served the UK Medical Research Council as a member of their Population and Systems Medicine Research Board, Training and Careers Group and Population Health Group, and the Welcome Trust as a member of their Advisory Group in Sustaining Health.