DIMACS Center
Rutgers University

Special Focus on Computational and Mathematical Epidemiology

Annual Report

August 2008

Participants who spent 160 hours or more:

Fred S. Roberts, Rutgers University, Principal Investigator
Martin Farach-Colton, Rutgers University
David Madigan, Rutgers University
S. Muthukrishnan, Rutgers University
Donald Hoover, Rutgers University
Nina Fefferman, Rutgers University
James Abello, Rutgers University

Other Participants:

Sunetra Gupta, Oxford University
David Krakauer, Santa Fe Institute
Simon Levin, Princeton University
Marc Lipsitch, Harvard School of Public Health
Ilya Muchnik, developing a cancer registry modeling project with James Abello and graduate student David Millman
Megan Murray, Harvard School of Public Health
David Ozonoff, Boston University
Burton Singer, Princeton University
Daniel Wartenberg, University of Medicine and Dentistry of New Jersey

Other Participants:

Workshop: Systems Biology of Infectious Diseases
Dates: August 13 - September 2, 2007

Organizers:
Charles DeLisi, Boston University
Simon Levin, Princeton University

DIMACS/DyDAn Workshop: Computational Methods for Dynamic Interaction Networks
Dates: September 24 - 25, 2007

Organizers:
Tanya Berger-Wolf, University of Illinois
Mark Goldberg, RPI
Malik Magdon-Ismail, RPI;
William "Al" Wallace, RPI

Workshop: Game Theoretic Approaches to Epidemiology and Ecology
Dates: October 15 - 17, 2007

Organizers:
Alison Galvani, Yale University
Tim Reluga, Los Alamos National Laboratory

Balancing Data Confidentiality and Data Quality: A two-day tutorial sponsored by DIMACS and

DyDAn
Dates: November 8 - 9, 2007

Organizer:
Larry Cox, CDC

Workshop: Ecologic Inference
Dates: November 28 - 30, 2007

Organizer:
Tom Webster, Boston University

DIMACS/DyDAn Workshop: Climate and Disease
Dates: April 7 - 8, 2008

Organizers:
Benjamin A. Cash, Center for Ocean-Land-Atmosphere Studies
Andrew Dobson, Princeton University
Jim Kinter, Center for Ocean-Land-Atmosphere Studies
Mercedes Pascual, University of Michigan

DIMACS/DyDAn Workshop: Investigation of Disease Clusters: Transitioning to the 21st Century and Beyond
Dates: May 6 - 8, 2008

Organizers:
Andrew Lawson, University of South Carolina
Daniel Wartenberg, Robert Wood Johnson Medical School

US-Africa Advanced Study Institute on Mathematical Modeling of Infectious Diseases in Africa
Dates: June 30 - July 11, 2008

Organizers:
Wayne Getz, UC Berkeley
Abba Gumel, University of Manitoba
Fritz Hahne, AIMS
John Hargrove, SACEMA
Simon Levin, Princeton University
Edward Lungu, University of Botswana;
Fred Roberts, DIMACS
Alex Welte, Wits University

Workshop: Modeling the Impact of Policy Options during Public Health Crises
Dates: July 27 - August 1, 2008

Organizers:
Fred Brauer, University of British Columbia
John W. Glasser, CDC
Zhilan Feng, Purdue University

Supplemental Workshop: Integrating Bayesian Regression into Blackbook, Learning with Emergent Classes and Higher-Order Links

Participants

- Fred Roberts (Rutgers)
- Paul Kantor (Rutgers)
- David Madigan (Columbia-Rutgers)
- William Pottenger (Rutgers)
- Tong Zhang (Rutgers)
- David Lewis (David Lewis Associates)
- Vladimir Menkov (Consultant - Programmer)
- Mark Dilsizian (Lehigh and Rutgers)

Graduate students:
Steve Bellan, PhD student in Envrionmental Science, Policy and Management at the Getz Lab at the University of California, Berkeley

Other Collaborators

Graham Cormode, Bell Laboratories and AT&T Laboratories, Co-Organizer of Tutorial on Data Mining and Epidemiology

Larry Cox, CDC, Co-Organizer, Working Group: Data De-Identification, Combinatorial Optimization, Graph Theory, and the Stat/OR Interface

Troy Day, Queens University, Co-Organizer, Workshop: Evolutionary Considerations in Vaccine Use

Ding-zhu Du, University of Texas at Dallas, Co-Organizer of Workshop: Combinatorial Group Testing

Alison Galvani, Yale University., Co-Organizer, Workshop: Evolutionary Considerations in Vaccine Use

Abba Gumel, University. of Manitoba, Co-Organizer, Workshop: Evolutionary Considerations in Vaccine Use

Frank Hwang, Chiatong University, Co-Organizer of Workshop: Combinatorial Group Testing

Ramanan Laxminarayan, Resources for the Future, Co-Organizer, Workshop: Economic Epidemiology

Randy Linder, University of Texas, Co-Organizer, Workshop and Working Group: Reticulated Evolution

Catherine Macken, Los Alamos National Labs, Co-Organizer, Workshop: The Epidemiology and Evolution of Influenza

Colleen Martin, CDC, Co-Organizer of Working Group on BioSurveillance Data Monitoring and Information Exchange

Bernard Moret, University of New Mexico, Co-Organizer, Workshop and Working Group: Reticulated Evolution

Alan Perelson, Los Alamos National Labs, Co-Organizer, Workshop: The Epidemiology and Evolution of Influenza

Allen Rodrigo, University Of Auckland, Co-Organizer, Workshop: Phylogenetic Trees and Rapidly Evolving Pathogens, Co-Organizer of Working Group Phylogenetic Trees and Rapidly Evolving Pathogens

Henry Rolka, CDC, Co-Organizer of Working Group on Adverse Event/Disease Reporting, Surveillance and Analysis; Co-Organizer of Working Group on BioSurveillance Data Monitoring and Information Exchange

Dave Smith, National Institute of Health (NIH), Co-Organizer, Workshop: Economic Epidemiology

Mike Steel, University of Canterbury, Co-Organizer, Workshop: Phylogenetic Trees and Rapidly Evolving Diseases, Co-Organizer, Working Group Meeting: Phylogenetic Trees and Rapidly Evolving Diseases

Katherine St. John, The City University of New York, Lehman College, Organizer of Tutorial: Phylogenetic Trees and Rapidly Evolving Pathogens

Claudio Struchiner, Oswaldo Cruz Foundation, Co-Organizer, Workshop: Evolutionary Considerations in Vaccine Use

Partner Organizations:

Princeton University: Collaborative Research; Personnel Exchanges. Partner organization of DIMACS. Individuals from the organization participated in the program planning, organization, and the research.

AT&T Labs - Research: Collaborative Research, Partner organization of DIMACS. Individuals from the organization participated in the program planning.

Telcordia Technologies: Collaborative Research; Personnel Exchanges. Partner organization of DIMACS. Individuals from the organization participated in the program planning.

NEC Laboratories America: Collaborative Research; Personnel Exchanges. Partner organization of DIMACS. Individuals from the organization participated in the program planning.

Lucent Technologies, Bell Labs: Collaborative Research; Personnel Exchanges. Partner organization of DIMACS. Individuals from the organization participated in the program planning, organization and research.

Office of Naval Research: Financial Support

Alfred P. Sloan Foundation: Financial Support

Burroughs-Wellcome Fund: Financial Support

New Jersey Commission on Science and Technology: Financial Support

IBM Research: Collaborative Research; Personnel Exchanges. Partner organization of DIMACS.

Individuals from the organization participated in the program planning.

Microsoft Research: Collaborative Research. Partner organization of DIMACS. Individuals from the organization participated in the program planning.

Avaya Labs: Collaborative Research. Partner organization of DIMACS. Individuals from the organization participated in the program planning.

American Statistical Association: Financial Support

Hewlett-Packard Labs: Collaborative Research. Partner organization of DIMACS. Individuals from the organization participated in the program planning.

Centers for Disease Control: Collaborative Research; Personnel Exchanges. Individuals from the organization participated in the program planning and working group/workshop organization.

Activities

This is a five-year special focus, which started in summer of 2002, following the design of our Center's pioneering special focus on Mathematical Support for Molecular Biology. In this special focus, the Center seeks to:

- Develop and strengthen collaborations and partnerships between mathematical scientists (mathematicians, computer scientists, operations researchers, statisticians) and biological scientists (biologists, epidemiologists, clinicians).

- Identify and explore issues in mathematics and computer science that need to be resolved to make progress on important problems in epidemiology.

- Identify and explore methods of mathematical science not yet widely used in studying problems of epidemiology and introduce epidemiologists to them - with an emphasis on methods of discrete mathematics (including discrete probability) and the algorithms, models, and concepts developed in the field of theoretical computer science.

- Introduce outstanding young people from both the mathematical/computer science and biological communities to the issues and problems and challenges of computational and mathematical epidemiology.

- Involve biological and mathematical scientists together to define the agenda and develop the tools of computational and mathematical epidemiology.

The special focus consists of a research program featuring “working groups” concentrating on specific research topics and a program integrating research and education through a series of workshops and tutorials. The 1994-2000 DIMACS Special Focus on Mathematical Support for Molecular Biology played a central role in laying the groundwork for the field of computational molecular biology, led many fledgling concepts and methods grounded in the mathematical sciences to become standard tools in the biological sciences, produced lasting partnerships between biological and mathematical scientists, and introduced many of today's leaders in computational biology to the field and to each other. This special focus has been doing the same.

Epidemic models of infectious diseases go back to Daniel Bernoulli's mathematical analysis of smallpox in 1760 and have been developed extensively since the early 1900s. Hundreds of mathematical models have been published since, exploring the effects of bacterial, parasitic, and viral pathogens on human populations. The results have highlighted and formalized such concepts as the notion of a core population in sexually transmitted diseases and made explicit other concepts such as herd immunity for vaccination policies. Relating to persistent infections, key pathogens that have been studied are: Malaria, *Neisseria gonorrheae*, *M. tuberculosis*, *HIV*, and *T. palladum*. Important issues such as drug-resistance, rate of spread of infection, epidemic trends, and the effects of treatment and vaccination all have been addressed through mathematical modeling approaches, which with the help of computational tools have provided new insights. Yet, for many infectious diseases, we are far from understanding the mechanisms of disease dynamics. The strength of the modeling process is that it can lend insight and clarification to existing data and theories. Mathematical models provide a unique approach to representing and studying the integrated behavior of complex biological systems and enable us to compare and contrast existing theories of the dynamic interactions in a complex system. The size of modern epidemiological problems and the large data sets that arise call out for the use of powerful computational methods for studying these large models. As pointed out by Levin, Grenfell, Hastings, and Perelson in a 1997 article in *Science*, “imaginative and efficient computational approaches are essential in dealing with the overwhelming complexity of [such] biological systems.” New computational methods are needed to deal with the dynamics of multiple interacting strains of viruses though the construction and simulation of dynamic models, the problems of spatial spread of disease through pattern analysis and simulation, and the optimization of drug design through hierarchical and other search methods on adaptive landscapes.

Statistical methods have long been used in mainstream epidemiology largely for the purpose of evaluating the role of chance and confounding associations. Considerable effort is expended by epidemiologists to ferret out sources of systematic error (“bias and confounding”) in the observations and to evaluate the role of uncontrollable error (using statistical methods) in producing the results. Interpretation of the results usually depends upon correlative information from the medical and biological sciences. The role of statistical methods in epidemiology is changing due to the large data sets that are arising and this calls for new methods and new approaches, making use of modern information technology for dealing with huge data sets of information on disease patterns.

A smaller but venerable tradition within epidemiology has considered the spread of infectious disease as a dynamical system and applied difference equations and differential equations to that end. But little systematic effort has been made to apply today's powerful computational methods to these dynamical systems models and few computer scientists have been involved in the process. We hope to change this situation. Probabilistic methods, in particular stochastic processes, have also played an important role. However, here again, few computer scientists have been involved in efforts to bring the power of modern computational methods to bear.

A variety of other potentially useful approaches to epidemiological issues have not yet attracted the attention of many in the computer science community nor are the methods made widely available to biological scientists. For example, many fields of science, and in particular molecular biology, have made extensive use of the methods and techniques of discrete mathematics (broadly defined), especially those that exploit the power of modern computational tools. These are guided by the algorithmic and modeling methods of theoretical computer science that make these tools more available than they have been in the past. Yet, these methods remain largely unused in epidemiology. One major development in epidemiology that makes the tools of discrete mathematics and theoretical computer science especially relevant is the use of Geographic Information Systems (GIS). These systems allow analytic approaches to spatial information not used previously. Another development is the availability of large and disparate computerized databases on subjects containing information on many attributes that might be related to disease status.

The role of discrete mathematics and theoretical computer science has also become important with the increasing emphasis in epidemiology of an evolutionary point of view. To fully understand issues such as immune responses of hosts; co-evolution of hosts, parasites, and vectors; drug response; and antibiotic resistance; among others, biologists are increasingly taking approaches that model the impact of mutation, selection, population structure, selective breeding, and genetic drift on the evolution of infectious organisms and their various hosts. Epidemiologists are only beginning to become aware of some of the computer science tools available to analyze these complex problems, such as methods of classification and phylogenetic tree reconstruction grounded in concepts and algorithms of discrete mathematics and theoretical computer science and developed in connection with the explosion in "computational biology," a field in which DIMACS has been a pioneer. Many of the recent methods of phylogenetic tree reconstruction resulted from the DIMACS Special Focus on Mathematical Support for Molecular Biology are described in the DIMACS website in the reports on the accomplishments of the earlier Special Focus. Yet, a great deal more needs to be done.

One important modern topic in theoretical computer science that arose in epidemiology is the theory of group testing, which arose in connection with testing millions of World War II military draftees for syphilis. The idea is to avoid testing each individual and instead to divide them into groups and determine if some individual in the group is positive for the disease, updating the process with groups that test positive. The modern theory of group testing is heavily influenced by combinatorial methods, in particular by the methods of combinatorial designs and coding theory, and many modern algorithmic methods, developed by theoretical computer scientists, are not yet widely known or used in epidemiology.

Mathematical methods of formal logic and ordered algebraic systems have been used to develop the foundations for a theory of measurement with important uses in the physical sciences and, more recently, in the social and biological. While this kind of measurement theory has been applied to data analysis in the social and biological sciences, it is virtually unknown in the epidemiology community (where the term "measurement theory" has other connotations), except to the extent that epidemiological studies use principles, grounded in but sometimes challenged by measurement theory, such as that certain kinds of statistical tests are inappropriate for ordinal data.

New interdisciplinary approaches, involving partnerships among mathematical scientists and biological scientists, epidemiologists, and clinicians, offer the promise for making progress on modern epidemiological problems and should take both fields of epidemiology and mathematics/computer science in new and fruitful directions. Mathematical and computational methods seem especially relevant in light of recent modeling approaches to emerging infectious diseases such as the vector-borne diseases from West Nile virus, Eastern equine encephalitis virus and Borrelia burgdorfei (Lyme disease); the spread of "mad cow" disease (transmissible spongiform encephalopathy; and HIV/AIDS. Control measures for these diseases often have counter-intuitive consequences only revealed after sophisticated mathematical analysis. Similar advances as a result of applications of mathematical and computational modeling have not been as evident in the area of chronic disease epidemiology, although work of considerable promise is being done, for example on modeling of the progression of cancer. In this special focus, we are considering both infectious and non-infectious diseases, and we will explore mathematical and computational approaches to animal and plant diseases as well as to human diseases.

The tutorials, workshops, and working group meetings that were held during this year are as follows:

Workshop: Systems Biology of Infectious Diseases

Dates: August 13 - September 2, 2007
Location: Aspen Center for Physics, Aspen, Colorado

Organizers: Charles DeLisi, Boston University; Simon Levin, Princeton University
Attendance: 23

The annual physics-astrophysics program at the Aspen Center for Physics was held from May 28 to September 2, 2007. The Center provided a place for physicists and astrophysicists to work on their research with minimal distraction in a stimulating atmosphere, and in a location of great natural beauty. Applications were welcome from any physicist or astrophysicist who had a serious program of research to be carried out at the Center. The Aspen Center for Physics is committed to a significant participation of women and under-represented groups in all of the Center's programs.

Workshops
Equally important to the Aspen Summer Program are the informal workshops that serve as focal points on topics of current interest. Workshops are very informal, with an extremely limited number of talks so that participants have ample time for informal discussion and to initiate new work. Information about the informal workshop on Systems Biology of Infectious Diseases scheduled for August 13 - September 2, 2007 is below:

The mathematical sciences have long spurred fundamental and applied advances in the life sciences. In the modern era a number of major transformations are associated with specific individuals: the profound physiological analyses of Helmholtz, and the Crick theory of X-ray scattering from double helices, are only two among many. More recently a major cultural shift toward mathematics and computation has been forced by the Human Genome Project which would not have been possible, absent the methods of mathematics and advanced computation.

The subject on which this workshop focused, infectious disease, is among the areas in which applied and fundamental quantitative science has played a major role for decades. Contributions ranged from differential equations models of disease dynamics in human populations--which provides, among other things, the basis for policy--to the quantitative tools of the civil engineer, which have increased life span by decades. More recently, advances in cell biology have transformed our understanding of disease related processes, including molecular changes that occur when pathogens infect cells; physical processes that underlie infection; and systemic changes in the host. We are now poised to understand the biological, chemical and physical determinants of host-to host transmission and changes in host range, and to begin integrating such understanding with changes in demographics, climate, globalization and so on. A central goal of infectious disease research is, therefore, integration across scales. This workshop offered a unique opportunity for physicists to exchange ideas with colleagues in the biomedical sciences on a range of topics relevant to emerging and remerging infectious disease, which cut across scales. These include the physical properties of proteins and nucleic acids and their interactions; the mechanisms governing the behavior of molecular motors; the physics of transcriptional and translational control; learning, memory and adaptation in cells, organs and individuals; and the dynamics of host-host and host parasite interactions.

DIMACS/DyDAn Workshop: Computational Methods for Dynamic Interaction Networks
Dates: September 24 - 25, 2007
Location: DIMACS Center, CoRE Building, Rutgers University
Organizers: Tanya Berger-Wolf, University of Illinois; Mark Goldberg, RPI; Malik Magdon-Ismail, RPI; William "Al" Wallace, RPI
Attendance: 63

A substantial body of research in various sciences aims at understanding the dynamics and patterns of interactions within populations, in particular how social groups arise and evolve. As a result of the advances in communications and computing technology, extreme amounts of data are being accumulated representing the evolution of large scale communication networks, such as the WWW, chatrooms, Blogs, and networks of bluetooth enabled handheld devices. Moreover, as small sensors become largely available and affordable, new research areas are exploiting the social networks resulting from those sensor networks data. Finding patterns of social interaction within a population has been addressed in a wide range applications including: disease modeling cultural and information transmission, intelligence and surveillance, business management, conservation biology and behavioral ecology.

The workshop focused on two complementary themes. On one hand it addressed the emerging importance of electronic communication networks, their social implications and how those facilitate the organization and coordination of activities of social groups. The second theme of the workshop was adapting and extending the computational methods developed in the context of communication and computer networks to the social interaction networks.

TOPICS:

- Modeling and simulation of dynamic social networks
- Measurement and comparison of dynamic social networks
- Community and social structure identification
- Identification of individual roles and behavioral patterns
- Visualization of large dynamic networks

Workshop: Game Theoretic Approaches to Epidemiology and Ecology

Dates: October 15 - 17, 2007
Location: DIMACS Center, CoRE Building, Rutgers University
Organizers: Alison Galvani, Yale University; Tim Reluga, Los Alamos National Laboratory
Attendance: 42

One of the challenges of public policy design is accounting for the inevitable conflicts between individual, economic and community interests. Such scenarios are common in public-health and environmental contexts, from the classic tragedy-of-the-commons dilemma to free-rider problems in preventative medicine. Policies that optimize for one set of interests are often ineffective because they rely on idealized assumptions about the behaviors of competing interests. Recent work has argued that a broader, multi-disciplinary approach to competing interests will lead to more robust public policies. This workshop will focus on the integration of epidemiology, behavioral science, economics, and mathematics to better understand these conflicts.

Balancing Data Confidentiality and Data Quality: A two-day tutorial sponsored by DIMACS and DyDAn

Dates: November 8 - 9, 2007
Location: DIMACS Center, CoRE Building, Rutgers University
Organizer: Larry Cox, CDC
Attendance: 21

Statistical summary data such as tabulations are built from data pertaining to individual entities (persons, households, businesses, organizations or groups). Statistical microdata are unit-record data containing multiple item responses pertaining to individual entities. Statistical data base query systems, once only a possibility, are becoming a reality. The need for data products that combine information across data bases and organizations is increasing and such data products arise in applications ranging over homeland security, health care, financial transactions, etc.

Typically the data from which these statistical data products are built is reported at the individual entity level and is confidential.

Ethical survey practice demands that confidential data pertaining to individual persons or entities not be revealed through released data products. Ethical concerns are often reinforced by legislation or regulation, such as the Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA) and the Health Insurance Portability and Accountability Act of 1996 (HIPAA). Confidentiality concerns have been addressed by researchers and government statisticians over several decades, resulting in a suite of increasingly sophisticated and effective methods for statistical disclosure limitation (SDL), several of which have been implemented in software and incorporated in the survey practices of government statistical agencies in the U.S. and abroad. Until very recently, however, the effects of disclosure limitation methods on data quality, completeness and usability have been largely ignored. The interplay between data confidentiality and data quality is a central subject of this tutorial.

This tutorial had three objectives: (1) to familiarize the student with statistical disclosure limitation and SDL methods; (2) to examine potential effects of SDL methods on data completeness, quality and usability; and, (3) to present SDL methods that, in addition to protecting confidentiality effectively, limit abbreviation or deterioration in the usability, quality and completeness of the released data product(s). Practical data quality questions include: What effect does the SDL method have on key statistics? What effect does the SDL method have on the distribution of the original data? How easy are disclosure-limited data to analyze compared to original data? Is analysis based on disclosure-limited data an acceptable substitute for analysis based on original data?

This tutorial covered the following topics: reasons for confidentiality protection; legal and regulatory requirements, including CIPSEA and HIPAA; legal and administrative solutions for restricting unauthorized access to confidential data; survey methods for restricting released data and for quantifying and limiting disclosure in tabulations, microdata and public use statistical data base query systems; using research data centers and controlled remote access to increase authorized access to confidential data; and, balancing the confidentiality protection provided by SDL methods with their effects on the usability, quality and completeness of released data products. Emphasis was placed on recognizing disclosure and evaluating the effectiveness of disclosure limitation strategies and their effects on data quality by means of lecture, discussion and simple numeric examples. Classroom notes, mathematical preliminaries, URLs, and references on disclosure limitation was provided. The tutorial was organized around types of data release-tabulations, microdata, data base query systems-but much of the material, particularly from the first day, is of general relevance.

Workshop: Ecologic Inference

Dates: November 28 - 30, 2007
Location: DIMACS Center, CoRE Building, Rutgers University
Organizer: Tom Webster, Boston University
Attendance: 19

In an individual-level study, epidemiologists collect information -- disease outcome, exposure and covariates -- for each subject. Ecologic studies, in contrast, collect and analyze aggregate data. As an example of the latter, one might compare average meat consumption with colorectal cancer by country. (The terms "ecologic" and "ecological" are used interchangeably in this context, but neither refers to ecology: They refer to "groups" or grouped data.) Ecologic studies are attractive because they are often inexpensive and easily conducted, relying on routinely collected data. However, ecologic studies are subject to very large potential biases: relationships between group-level variables do not necessarily reflect relationships between individual-level variables and vice versa. Furthermore, many nominally individual-level studies employ one or more group-level variables, raising the potential of ecologic bias in such partially-ecologic studies. While research has traditionally focused on bias arising from estimating individual-level parameters from fully aggregated data, other study designs are receiving increased attention: partially-ecologic studies and studies with contextual (purely group-level) effects. The workshop focused particularly on methods that combined individual and group level data. While most research on ecologic bias is theoretical, insight into the amount of ecologic bias actually occurring in real studies can be obtained from parallel analyses of the same data set on the individual and group levels. Both theoretical and applied papers are of interest. The workshop brought together epidemiologists and biostatisticians/mathematicians to discuss these developments. For additional background on ecologic bias, please see: http://www.cireeh.org/pmwiki.php/Main/Ecologicstudies

DIMACS/DyDAn Workshop: Climate and Disease

Dates: April 7 - 8, 2008
Location: DIMACS Center, CoRE Building, Rutgers University
Organizers: Benjamin A. Cash, Center for Ocean-Land-Atmosphere Studies; Andrew Dobson, Princeton University; Jim Kinter, Center for Ocean-Land-Atmosphere Studies; and Mercedes Pascual, University of Michigan
Attendance: 48

The latest IPCC report has strongly confirmed that the climate is changing. The present and future impact of climate change on infectious disease dynamics remains an important, but still controversial subject. Malaria is a major example of a public health burden around the tropics with the potential to significantly worsen in response to climate change; temperature as a limiting factor for the pathogen, and temperature and rainfall play a crucial role in determining the population dynamics of its mosquito vector. Similar concerns apply to other vector-borne and water-borne diseases, particularly given the existing evidence for the role played by climate at seasonal and inter-annual time scales (e.g. ENSO). This workshop explored the boundaries of current knowledge and in particular examined what role theory and mathematical (epidemiological) models can play in advancing the understanding and prediction of the coupling between two highly nonlinear phenomena: climate and infectious disease dynamics.

While dynamical models have contributed significantly to our general understanding of the population dynamics of infectious diseases, most often their application has been to assume stationary environmental conditions and examine the constraints that climate places on the geographical distribution of vector borne diseases. However, most infectious diseases are highly dynamic with outbreaks that vary in size from year to year, including intermittent epidemics, the emergence of new pathogens, and exacerbation in the prevalence of old ones. Thus, there is an urgent need to better understand the dynamics of infectious disease outbreaks and their response to climate variability; this needs to be developed in the context of longer-term environmental change.

The workshop was organized around the following main topics: (1) the consideration of dynamical approaches, vs. static ones, to address the effects of climate change in both time and space; (2) the relevant spatial and temporal scales of coupling of climate and epidemiological models for constructing both early-warning systems and future scenarios; and (3) the application of mathematical models and theory to address the synergy between climate change and other different aspects of human-induced change: the evolution of drug resistance, changing patterns of land-use, and socio-economic conditions.

The understanding and forecasting of these patterns requires an ecological perspective that builds on the long-history of mathematical models for infectious disease dynamics, but expands these efforts in several important ways to develop our ability to: (1) interface these models with existing data on both disease and environment, (2) consider different temporal scales of change to address the dynamics of outbreaks together with longer-term trends, and (3) bridge the different organizational scales of within and between host dynamics.

DIMACS/DyDAn Workshop: Investigation of Disease Clusters: Transitioning to the 21st Century and Beyond

Dates: May 6 - 8, 2008
Location: DIMACS Center, CoRE Building, Rutgers University
Organizers: Andrew Lawson, University of South Carolina, and Daniel Wartenberg, Robert Wood Johnson Medical School
Attendance: 38

Disease clusters, defined as local excesses of disease in space, time or space and time, represent an important but vexing problem in public health. Clusters are usually identified by community residents who believe that some unusual circumstance leading to unexpected illness has befallen their families, friends or neighbors. Clusters of leukemia are reported most often, although clusters of other cancers, birth defects and other adverse health outcomes are also reported. While there are many protocols to assess whether a given cluster is etiologic, i.e., due to an identifiable cause, there is no clear consensus about how best to conduct an investigation and reach a scientifically valid conclusion. A variety of statistical issues confront investigators of clusters. For example, since cluster reports typically are based on a handful of cases, it is possible that the observed excess is simply due to random variation, particularly if the investigator fails to adjust for multiple comparisons. On the other hand, the statistical power of traditional cluster analysis methods is fairly low, likely resulting in many false negatives which might cause investigators to miss true, etiologic clusters. In addition, assumptions are made about the amount of disease expected because large data sets are generally not available at a scale that would enable investigators to determine background rates of disease, such as at the census block, census tract or zip code. Some recent methods have begun to look at approaches for conducting prospective surveillance by analyzing data collected for each time unit (e.g., year) it is collected. These methods offer the opportunity the overcome some of the statistical limitations of traditional cluster analyses and provide a more appropriate perspective for health officials to use in responding to community concerns. The workshop brought together mathematicians, biostatisticians, epidemiologists and public health officials to develop an approach that, while statistically rigorous, is able to address the concerns of the public.

US-Africa Advanced Study Institute on Mathematical Modeling of Infectious Diseases in Africa
Dates: June 30 - July 11, 2008
Location: AIMS, Muizenberg, South Africa
Organizers: Wayne Getz, UC Berkeley; Abba Gumel, University of Manitoba; Fritz Hahne, AIMS; John Hargrove, SACEMA; Simon Levin, Princeton University; Edward Lungu, University of Botswana; Fred Roberts, DIMACS; and Alex Welte, Wits University University
Attendance: 24

Workshop: Modeling the Impact of Policy Options during Public Health Crises
Dates: July 27 - August 1, 2008
Location: Banff International Research Station
Organizers: Fred Brauer, University of British Columbia; John W. Glasser, CDC; and Zhilan Feng, Purdue University
Attendance: 44

We organized a Mathematical Epidemiology workshop, jointly sponsored by the Pacific Institute of Mathematical Sciences (PIMS) and Center for Discrete Mathematics and Theoretical Computer Science (DIMACS), at the Banff International Research Station (BIRS) during the summer of 2007. While a similarly entitled 2005 workshop at the BIRS brought together epidemiologists and mathematicians who use existing approaches to solve public health problems or develop new ones, a crucial ingredient was missing. We proposed to mix influential policymakers with able mathematicians and epidemiologists, particularly ones interested in collaborating.

The recent progress mentioned above has made the potential for collaboration between modelers and policymakers even more apparent to the organizers. But unless modelers fully appreciate the political challenges of policymaking and policymakers appreciate why models must be as simple as possible, but no more so, opportunities to develop sound public policy in preparation for or response to future health crises ? be they mutation of an avian influenza virus, increasing transmissibility person-to-person, and resulting in the next pandemic, or whatever ? may never materialize.
We believe a workshop in a setting conducive to discussion of the benefits of collaboration would repay dividends during future public health crises. Accordingly, we proposed to mix well-positioned members of both groups willing to work together, but possibly not fully appreciative of the benefits to themselves much less society, with ones who have not only weathered health crises, but learned what makes fruitful collaborations. We are confident that interactions between experienced and willing collaborators in this setting will lead to significant partnerships between modelers and policymakers.

We plan to organize sessions around modes of transmission, respiratory (e.g., avian influenza, severe acute respiratory syndrome, SARS), sexual (e.g., acquired immunodeficiency syndrome and cervical cancer due to the human immunodeficiency and papilloma viruses), vector-borne (e.g., dengue, equine encephalitis, malaria, West Nile), and so on, for each of which we will identify responsible policymakers, knowledgeable medical epidemiologists, and experienced modelers. Sessions were followed by discussions in informal settings facilitated by rapporteurs whose responsibilities include identifying means of furthering collaborations in their respective fields and ensuring they come to fruition. Thus, this workshop's product was a coherent plan of action versus the usual summary of discussions.

Supplemental Workshop: Integrating Bayesian Regression into Blackbook, Learning with Emergent Classes and Higher-Order Links

Dates:
Location:
Organizer:
Attendance: 8

During this year, we initiated work on a project on "Integrating Bayesian Regression into Blackbook, Learning with Emergent Classes and Higher-Order Links" that was funded by the intelligence community through their KDD program. The work centers around learning patterns from (often sparse and expensive) bodies of analyzed materials. The problems are conceptually quite distinct, although several of our solutions make use of, and extend the power of, the Bayesian Regression software (BXR) that we have developed in earlier intelligence-community-sponsored work. The primary task is to integrate BXR more fully with Blackbook and in particular tightly couple the two systems' semantic models. New work will extend BXR to deal with new training data and new class definitions in real time and to in general produce a more stream- oriented API for BXR; and will develop methods for learning patterns that involve higher-order relationships among real or aliased entities. We will also integrate the new work into Blackbook.

Integration into Blackbook involves data modeling (understanding approaches used by Blackbook, developing conventions for communicating this to BXR, and defining mappings back from BXR to the Blackbook semantic model. It also involves data formats and communication, system architecture, and data management. Our software is being developed so that it is ready to accept information from Blackbook as an interface to the analyst, and we are making modifications that will greatly simplify the process of computing the data for higher order paths. In connection with this work, we sent one of our graduate students, Mark Dilsizian, to take a course on Semantic Web w/RDF and OWL, given at the University of Texas-Dallas, and run by developers of Blackbook.

We are also working on a task we call "learning with emergent classes and training data." When information streams are processed over time, both the properties of the data and the classes of interest may change. We are extending our algorithms and software to situations where new training data, and new class distinctions to make, are provided incrementally. We are developing online learning algorithms that leverage both changes in stream properties and changes in discrimination sets to improve effectiveness. We are developing and implementing, in BXRtrain, an algorithm for learning from data streams with multiple and emerging classes, while supporting BXRtrain's Bayesian priors.

A final task is concerned with "higher-order path analysis." In addition to knowing how many agents there are at play, and knowing which of them are responsible for which particular products or activities, it also extremely important to know the links and relationships among these agents. One way to characterize the topology of the relationships in a graph is based on higher-order path analysis techniques. Thus, the strength of connections between entities reflects not only the most direct path, but indirect paths as well. We are developing ways to apply unsupervised higher-order path methods to measure the strength of links between entities/agents and integrating higher-order path statistics into BXR.

Findings

Numerous new results have occurred as a result of the workshops on Mathematical Epidemiology. Some of the examples follow:

1. Five arenaviruses are associated with rodent-transmitted diseases in humans. Banerjee and Allen have modeled the Machupo virus, also known as Bolivian hemorrhagic fever (BHF) in humans. It is spread through a rodent host Calomys callosus in the rodent population and on to humans. Three different models were analyzed, the first where all infected rodents are immunocompetent and recover, a second where all infected rodents are immunotolerant, and a third where the sexes are not differentiated. It is shown that for the second model the rodent infectiousness becomes asymptotically stable, and bistability can occur; and for the third model that two endemic equilibria occur, one of which is globally asymptotically stable and bistability does not occur.

2. Steve Bellan, a PhD student in Envrionmental Science, Policy and Management at the Getz Lab at the University of California, Berkeley created a model of arboviral transmission, which demonstrated that in light of recent evidence for age-dependent mosquito mortality, adult mosquito control is much less efficient compared to larval mosquito control than previously thought and that this result is magnified for arboviruses with short extrinsic incubation periods such as the chikungunya virus.

3. Economic Epidemiology – Nina Fefferman, working with one of her REU students from last summer, and in collaboration with Alison Galvani (Yale Univ.), have formulated a game theoretic model of the trade-offs involved in household economics, providing home care for sick children, children’s health individually, and a school’s interest in preventing transmission of disease among children. Preliminary results from this work were presented at the DIMACS Workshop on Game Theory and Epidemiology, and two papers resulting from these collaborations are currently in preparation.

4. Fundamental formulations of conceptual Privacy: Dr. Aaron Jaggard (at DIMACS) and Dr. Nina Fefferman have begun formulating a framework by which to rigorously analyze colloquial and field-specific concepts of privacy to facilitate cross-field communication. While this is mostly

separate from Epidemiological foci, one of our main points of concern is making rigorous and clear the expectations and needs for public health privacy in an otherwise rule-based, but opaque area of medical communication.

5. Dr. Jaques Kibambe from University of Pretoria and Dr. Nina Fefferman of DIMACS formulated a human capital model to investigate the economic trade-offs within a family for treatment and prevention of infectious disease among household members based on their contribution to family income and presumed immunocompetence (determined by age). They then expanded these models to investigate multi-generational effects, looking at how to provide appropriate incentives on a community level to allow individual families to minimize their exposure to both economic and health risks. This work was presented internally at a DIMACS Epi Seminar, but was also presented by Dr. Kibambe at his home institution's seminar series. Preparation of these results for publication has begun.

6. A new initiative on **Climate Change and Health** is an offshoot of the epidemiology project and is funded by Rutgers "Academic Excellence Fund". It relates to the Climate and Disease workshop held this year, and was a major topic of discussion at the Banff workshop.

African Initiative. You can throw in a paragraph on our successful African project in June 2007 and follow up in July 2008 -- advanced study institute for US and African grad students (with lots of minorities from US involved) and workshop. Gene is doing the report. Also, major new African initiative planned. I think you have a summary of the planned activities.

Training & Development Section

All of the Workshops have opened doors for researchers and students to work collaboratively together. This extends well beyond the participants at a specific workshop. For example, Dr. Louise Russell, a Research Professor at the Institute for Health at Rutgers University, speaks of her interactions with Dr. Alison Galvani, organizer of the Workshop on Game Theoretic Approaches to Epidemiology and Ecology, which led to a graduate student in Economics, whom she knew quite well, securing a postdoctoral appointment in Professor Galavani's laboratory.

A second example is of an attendee at the Workshop on Immuno-Epidemiology, Dr. Linda Allen from Texas Tech University, who used the knowledge gained at that workshop to guide a Master of Science thesis along the same lines. The resulting publications have further extended the reach of the workshop. (See the publication list for Banerjee.)

Dr. Mick Roberts, Director of the Centre for Mathematical Biology at Massey University, Auckland, New Zealand, presented some results from joint research with Oxford Univesity, on HIV/AIDS and pandemic influenza at two DIMACS Epidemiology workshops. This work has resulted in two graduate student projects and future publications.

Steve Bellan, Ph.D student in Environmental Science, Policy and Management at UC Berkeley describes his grow through participation in the workshops by saying the following: "This work has led me to think about vector-borne diseases more thoroughly and led to conversations with colleagues on the importance of mosquito demograph in vector-borne disease models. At this workshop and through the additional work I have become increasingly aware of the disconnect between mathematical epidemiology, classical epidemiology, and applied disease control. I am intrigued by the possibility of connecting these three fields through research and teaching."

Outreach Activities

This Special Focus is closely intertwined with our Center's efforts to link mathematics and computer science with biology in the high schools. Project participants have been working with the Summer 2007 DIMACS Bio-Math Connection (BMC), which is aimed at introducing high school math, computer science, and biology teachers to topics at the interface. This project is informing the BMC effort and specific topics from the project are being adapted for use in BMC. The materials developed by BMC participants will consist of modules that can be flexibly adapted for use in a variety of courses at a variety of grade levels in both biology and mathematics. The project is run by DIMACS in collaboration with the Consortium for Mathematics and its Applications (COMAP) and Colorado State University. More information on the Field Testers workshop, that just ended, is available at http://dimacs.rutgers.edu/BMC/FieldTesters/2007/index.html Roughly twenty teachers were trained to use three modules integrating math and biology – Spider Silk, Genetic Inversion, and Biomatrices.

Researchers from Howard University and Morgan State University are spending the summer at DIMACS working on "Potential uses of entropy in biosurveillance" through a Department of Homeland Security-sponsored program aimed at faculty and students from minority-serving institutions. Nina Fefferman (DIMACS) and a team of two professors, two graduate students, and two undergraduate students are looking into coding theory techniques that can be applied to epidemiology. They hypothesize that the signal-to-noise ratio in reported disease incidence data may increase during the occurrence of disease outbreaks and that these increases may be observable when analyzed using entropy measures. If the hypothesized increases are statistically quantifiable, and if they are found to occur early enough in the progression of disease throughout the affected population, this may prove an invaluable tool for early-warning biosurveillance. This activity has been closely coordinated with and is a clear offshoot of our NSF grant.

Under other funding, but coordinated with and a clear offshoot of this grant, DIMACS and the South African Centre for Epidemiological Modeling and Analysis (SACEMA) held a 3-day workshop on mathematical modeling and infectious diseases in Africa in September 2006. Then in June 2007, DIMACS/SACEMA together with the African Institute for Mathematical Sciences (AIMS), held a two week short course and a subsequent 3-day "capstone" workshop. The short course was aimed at training United States and African graduate students and postdoctoral fellows in mathematical epidemiology and the control of emerging and re-emerging infectious diseases. The capstone workshop served as a culmination to the short course and helped in furthering research and forming collaborations on the modeling of infectious diseases in Africa.

Outreach to numerous communities has continued. Dr. Arnie Levine hosted a meeting at the Institute for Advanced Study, organized by Dr. Raul Rabadan, which brought tougher a number of the participants from the Workshop on Systems Biology of Infectious Diseases, and others to talk about open problems and collaborative work on flu.

Books

S. Ji, A. Chaovalitwongse, N. Feferman, W. Yoo, and J. E. Perez-Ortin, "Mechanism-based clustering of genome-wide transcript levels: roles of transcription and transcript-degradation rates," *Clustering Problems in Biological Networks*, ed. A. Chaovalitwongse (to appear).

X. Xue and D. R. Hoover, "Statistical methods in cancer epidemiological studies," Cancer Epidemiology 1: (2008).

Papers

J. Abello and M. Capalbo, "An approximation algorithm to the modified quarantine problem on expander graphs" (submitted).

J. Abello and M. Capalbo, "Blocking sequences in infinite grids" (submitted).

J. Abello, and M. Capalbo, "Max cliques in sparse power law graphs with large clustering coefficients," (submitted).

J. Abello, B. Gaudin, H. J. Schulz, C, Tominski, "Name that cluster: text vs graphics" (submitted).

C. Banerjee, "The dynamics of mathematical models for Machupo viral infection in rodent populations". M.S. thesis at Texas Tech University, Lubbock, TX USA, December 2007.

C. Banerjee, L.J.S. Allen, and J. Salazar-Bravo. "Models for an arenavirus infection in a rodent population: consequences of horizontal, vertical, and sexual transmissions". *Mathematical Biosciences and Engineering* (accepted for publication).

S. Bellan, "Reevaluating vector control in the context of age-dependent mosquito mortality: Consequences for the control of chikungunya". CDC's *Emerging Infectious Diseases.* (submitted May 2008).

B. Dembele, A. Friedman, and A. -A. Yakubu, "Malaria model with periodic mosquito birth rate," *Mathematical Biosciences* (submitted).

B. Dembele, A. Friedman, and A. -A. Yakubu, "Malaria model on the impact of drug administration protocols" (in preparation).

P.A. Dreyer, Jr., and F.S. Roberts, "Irreversible *k*-threshold processes: Graph-theoretical threshold models of the spread of disease and of opinion" (in preparation).

N. H. Fefferman and K. L. Ng, "Species-specific behavior affects disease spread throughout and ecosystem" (submitted).

M. Hauben, D. Madigan, S. Reisinger, A. Hochberg, and D. O'Hara, "Data mining in pharmacovigilence: computational cost as a neglected performance parameter," *Drug Safety*, International Journal of Pharmaceutical Medicine, 21 (5):319-323 (2007).

P.A. Dreyer, Jr. and F.S. Roberts, Irreversible k-threshold processes:

Graph-theoretical threshold models of the spread of disease and of
opinion, Discrete Applied Math., (to appear).

F.S. Roberts, Applications of Measurement Theory/Meaningfulness in
Epidemiology/Public Health, (in prearation).

Q. Xia and D. R. Hoover, "A procedure for group sequential comparative Poisson trials," *Biopharmaceutical Statistics*, 17 (5):869-881 (2007).

Reports

Nina Fefferman is editing the publication of a special issue of the journal *Annales Zoologici Fennici* dedicated to the uses of simulation modeling techniques in the investigation of questions in biology and medicine. The issue "Biological Experimentaion *in silico*", was due for publication last year, but had some problems with timing and is therefore is now scheduled for publication this year.

Talks

F. Roberts, "Climate and Disease", at Modeling the Impact of Policy Options during Public Health Crises, Banff International Research Station, July 2008)

F. Roberts, " How Modelers can Help Policymakers before and during Health Crises (The Case of TOPOFF 3)", at Modeling the Impact of Policy Options during Public Health Crises, Banff International Research Station, July 2008)

F. Roberts, "Applications of Measurement Theory/Meaningfulness inEpidemiology/Public Health", Plenary talk at European Mathematical Psychology Group meeting,Luxembourg, September 2007.

F. Roberts, "Applications of Measurement Theory/Meaningfulness in Epidemiology/Public Health", Plenary talk at MINET – Europeanm Commission Measuring the Impossible Network Training Course, Genoa, Italy, June 2008.

F. Roberts, "Why Bio-Math? Why Now?", at DIMACS BioMath Connection Field Testers Workshop, July 2007 and July 2008.

F. Roberts, "Meaningless Statements in Epidemiology",

Main Web Site

http://dimacs.rutgers.edu/SpecialYears/2002_Epid/

Other Specific Products

Web Pages

Web page for the workshop on “Systems Biology of Infectious Diseases.”
http://dimacs.rutgers.edu/Workshops/InfectiousDisease/

Web page for the workshop on “Computational Methods for Dynamic Interaction Networks.”
http://dimacs.rutgers.edu/Workshops/Dynamic/

Web page for the workshop on “Game Theoretic Approaches to Epidemiology and Ecology .”
http://dimacs.rutgers.edu/Workshops/Ecology/

Web page for the workshop on “Balancing Data Confidentiality and Data Quality: A two-day tutorial sponsored by DIMACS and DyDAn.”
http://dimacs.rutgers.edu/Workshops/DataQuality/ (Second Meeting)

Web page for the workshop on “Ecologic Inference.”
http://dimacs.rutgers.edu/Workshops/Ecologic/

Web page for the workshop on “Climate and Disease.”
http://dimacs.rutgers.edu/Workshops/Climate/

Web page for the workshop on “Investigation of Disease Clusters: Transitioning to the 21st Century and Beyond”
http://dimacs.rutgers.edu/Workshops/Clusters/

Web page for the workshop on “US-Africa Advanced Study Institute on Mathematical Modeling of Infectious Diseases in Africa.”
http://dimacs.rutgers.edu/Workshops/AIMS2008/

Web page for the workshop on “Modeling the Impact of Policy Options during Public Health Crises.”
http://dimacs.rutgers.edu/Workshops/BIRS/

Software

In 2004 and 2005 DIMACS, through this project, fostered a collaboration between David Ozonoff (Boston University) and Alex Pogel (U. of New Mexico) that has since flourished. During 2006, Pogel's work on Ozonoff's SuperFund grant has seen much progress made toward building and distributing an open source concept lattice analysis tool for epidemiologists, which will generate contingency tables from concept lattices with a few clicks on lattice nodes. This work stemmed originally from this NSF project. Pogel has hired a team of software engineers, written a specification document, and has a well-defined five-year plan for the completion of the project. The first version of this software is expected to appear at SourceForge.net in mid-2008, and the product will be publicized in the 2009-2010 academic year. Also, in Pogel's work at PSL, various features described in the Ozonoff, Pogel, and Hannan contribution to the AMS-DIMACS Special Volume on Computational Epidemiology (“Generalized Contingency Tables”) have been incorporated into Seqer (pronounced “seeker”), the data analysis software developed by Pogel's group over the past three years.

Contributions

Contributions within Discipline

This Special Focus is by nature multi-disciplinary. It involves applying methods of computer science, statistics, and mathematics to the problems in epidemiology. Introducing people to this combination of disciplines has been a key goal and a key accomplishment of this project. Below is a selection of comments from project participants indicating their assessment of the impact of this grant.

Kaus Dietz, Director of the Department of Medical Biometry at the Eberhard-Karls-University of Tübingen in Germany said "In December 2006 I attended the DIMACS-Workshop on Immuno-epidemiology. I met for the first time Professor Sungchul Ji, Department of Pharmacology and Toxicology, Rutgers University. We started a collaboration on the analysis of genome-wide mRNA levels measured during diauxic shift in yeast cells."

Sungchul Ji from the Department of Pharmacology and Toxicology at Rutgers University said "I want to thank you and your colleagues for organizing the above workshop this week. I learned a lot from the workshop, the first such meeting I have attended in immuno-epidemiology. The meeting opened my eyes to the field, which I may enter in the future as a theoretical cell biologist. One concrete result of my attending the workshop was meeting Klaus Dietz whom Hans Heesterbeek introduced to me. During the meeting Klaus and I began to analyze some of the budding yeast microarray data that I have been working on during the past two years. After taking Kalus to the Newark Airport on the last day, Klaus continued his calculations with his laptop at a bar in the airport and found out (about 30 minutes before his flight time) that my data are lognormally distributed. To me this is a totally unexpected finding and prompted me to write the attached email to ask him to expand his calculations. If Klaus finds lognormal distributions for *certain* variables we will have (I believe) results novel and important enough to construct a short communication to a journal such as the Journal of Theoretical Biology. Again, many thanks for all the good work that you and your colleagues at DIMACS have been doing over the years. I have been benefiting from these workshops enormously, but have been slow in acknowledging my indebtedness to DIMACS until now."

Gavin Welch from the Epidemiology and Biometry Core at Eastern Virginia Medical School said "Thanks for a wonderful series of workshops. As a direct result of the four DIMACS Comp-Epi workshops I attended: (1) One of the participants at the host-pathogen coevolution workshop gave a seminar on his influenza modeling work at my institution. We invited him because I learned of his work at the workshop. (2) I've used several examples in my classes that I've learned about by talking with other participants at DIMACS workshops. Most of the examples came from discussions over lunch."

Abdul-Aziz Yakubu from Howard University said "Participation resulted in collaborations with Avner Friedman and my student, Bassidy Dembele. We're working on Malaria epidemic in Mali. To date, we've submitted a paper for publication. The title of the paper is 'Malaria model with periodic mosquito birth rate.' We're currently working on a second paper on the Impact of Malaria Drug Administration in Mali."

One of the biggest contributions of this project has been the development of a new initiative to bring together scientists from the United States and various African countries to collaborate on mathematical modeling of infectious diseases in Africa. One specific DIMACS workshop was instrumental in leading to the development of the proposed activity in Africa. This was a workshop on "Evolutionary Aspects of Vaccine Use," organized at DIMACS at Rutgers University in June 2005, and stemming from the work of the DIMACS working group on Methodologies for Comparing Vaccination Strategies. The workshop, organized by Troy Day (Queen's University, Ontario), Alison Galvani (Yale), Abba Gumel (U. of Manitoba), and Claudio Struchiner (Oswaldo Cruz Foundation, Brazil), examined general evolution-related questions for diseases for which there is a vaccine or hope for one. There were 30 scientists participating, from diverse backgrounds, including mathematicians, epidemiologists, virologists, immunologists, and vaccine developers. The special problems of vaccination strategies in Africa that arose in this workshop were one of the primary motivations that led Abba Gumel to propose that DIMACS sponsor a short course and workshop that directly focuses on mathematical modeling of infectious diseases in Africa.

Following that workshop, DIMACS and the South African Centre for Epidemiological Modeling and Analysis (SACEMA) held a 3-day Workshop on mathematical modeling and infectious diseases in Africa, at the School of Computational and Applied Mathematics at the University of the Witwatersrand, Johannesburg, South Africa, on September 26-28, 2006. The workshop brought together scientists from the United States and various African countries, as well as junior researchers and students and provided an agenda for future collaborations. It exposed junior United States scientists and students to the special challenges of modeling the spread of disease in Africa and the opportunities to collaborate with Africans in developing and applying the tools of mathematical modeling to the tremendous health problems caused by such diseases as HIV/AIDS, tuberculosis and malaria, as well as the possible interventions in the case of major new health threats such as pandemic influenza in an underdeveloped region of the world.

Subsequently, DIMACS, SACEMA, and the African Institute for Mathematical Sciences (AIMS), developed plans for a two week short course on mathematical modeling and infectious diseases in Africa. The short course, held at AIMS in Cape Town, South Africa, on June 11-22, 2007, trained junior United States and African graduate students and postdoctoral fellows in mathematical epidemiology and the control of emerging and re-emerging infectious diseases. Those who completed the short course participated in a 3-day “capstone” workshop designed to establish collaborations. The capstone workshop was held at the new conference facilities of the Stellenbosch Institute for Advanced Study (STIAS) on June 25-27.

Long term DIMACS visitor Nina Fefferman said “Throughout the past year, I have continued my research into the varying success of organizational strategies of worker allocation under conditions of disease related workforce depletion. The results of these ongoing endeavors have been presented at two DIMACS workshops. Additionally, work to extend their applicability, incorporating economic modeling, is now underway in collaboration with researcher Dr. Ramanan Laxminarayan (of Resources for the Future), whom I met at a DIMACS workshop in 2005. Further possibilities for extensions of this work to examine alternate scenarios of economic impact are just beginning in collaboration with Dr. Martin Meltzer (of the CDC), whom I met at the recent DIMACS Capstone Workshop in South Africa.

“Due to our mutual involvement with DIMACS, Prof. Dina Fonseca (at the Rutgers Univ. Center for Vector Biology) and I met to discuss our mutual interest in temporal constraints on mosquito reproduction. As an outgrowth of this meeting, we have begun a collaboration to investigate the impact of mosquito habitat selection on invasive spread of a newly introduced mosquito in the state of NJ. Once accomplished, this work will have profound impact on the capability for predictive modeling of mosquito-borne diseases as this species is a known disease vector in its native habitats. This work involved the recruitment of three other researchers, Prof. Rick Lathrop, Prof. Jim Miller, and Prof. Randy Gaugler (of the Rutgers Univ. Depts. of Ecology and Evolution, Oceanography and Marine Science, and the Center for Vector Biology, respectively). Together, the five of us have now submitted a proposal for funding of this work to the NSF Directorate of Biological Sciences.

“As the result of a conversation had during a dinner at the DIMACS workshop on The Epidemiology and Evolution of Influenza, I have organized and am editing the publication of a special issue of the journal *Annales Zoologici Fennici* dedicated to the uses of simulation modeling techniques in the investigation of questions in biology and medicine. This issue is due for publication in October of this year. This issue would never have been organized had it not been for the interest and encouragement expressed at that DIMACS workshop dinner.

“All of these projects and collaborations were made possible by my involvement with DIMACS over the past year. I feel that my perspective, even of my own, individual research has been broadened substantially by this opportunity and I am very grateful. I feel lucky to have been a part of this focus in Computational and Mathematical Epidemiology and am looking forward to continuing with these projects and these people.”

Sungchul Ji describes his collaboration with Nina Fefferman as follows: “Dr. Fefferman attended one of my seminars given at the DIMACS Center at Rutgers on April 17, 2006 invited by James Abello. Based on the content of my seminar entitled ‘The Simpson-Elsasser-Wolfram (SEW) Framework for Modeling the Living Cell,’ Nina asked me to contribute a paper to a special edition of the journal Ann. Zool. Fennici, for which she is an editor. This led to writing a set of three manuscripts applying the network theory to cell biology. It was through the influence of James Abello, Nina Fefferman and Art Chaovalitwongse whom I met through DIMACS Workshops and Seminars that I was motivated to learn and apply network concepts and theories to molecular and cell biology. In the process, I have been to led to formulate what appears to be a novel theory of biological networks and apply it to modeling the living cell.”

Suzanne Lenhart from the University of Tennessee said “I started a new collaboration with Edward Lungu and Abba Gumel and will report back on the progress of that work about HIV epidemic models including the educational information impact.” Abba Gumel is from the University of Manitoba in Winnipeg, Canada and Edward Lungu is from the University of Botswana, in Gaborone, Botswana.

Jonathan Read from the Mathematics Institute at the University of Warwick collaborated with Nina Fefferman (DIMACS) on a variety of projects while visiting DIMACS. The primary focus of their work was expanding the research that Fefferman and Kah Loon Ng began to examine under what general rules self-organizing dynamic social networks formed similar convergent structures using centrality measures as a metric for similarity. This work continued at the DIMACS funded workshop in Edinburgh. Second, Read collaborated with Fefferman to design a novel presentational encoding strategy for social network graphs which will allow a rapid characterization of the potential for the spread of disease and/or information within the network depending on the individual through which occurs the primary introduction. Third, Read and Fefferman collaborated on models examining the difference in selective pressures acting on pathogen evolution caused by physiological constraints from host-immune evasion. Lastly, Read presented his own research into the social network structure of academia and during his visit collaborated with James Abello to create a clearer visual representation of his results. Together they were able to construct an image which clearly communicated greater levels of complexity in the network than had been previously possible.

Regarding Jonathan Read, Nina Fefferman said: “During the past year, I met researcher Jonathan Read at the DIMACS workshop on Models of Co-Evolution of Hosts and Pathogens. Together, we are working to develop models of the impact of intra-host immunology and physiology on the selective pressures driving pathogen virulence. This work is ongoing. We also collaborated to expand some of the work done by Fefferman & Ng (see above), attempting to determine under what general rules self-organizing dynamic social networks formed similar convergent structures using centrality measures as a metric for similarity. This work is also ongoing. In addition to those models, I have continued collaborating with Read to design a novel presentational encoding strategy for social network graphs which will allow a rapid characterization of the potential for the spread of disease and/or information within the network depending on the individual through which occurs the primary introduction. This work is ongoing and was continued in person after our initial meeting at the October 2006 workshop during Read’s brief visit to DIMACS in March of 2007, and subsequently also at the DIMACS workshop in May 2007 on Spatio-temporal and Network Modelling of Diseases. Each of these DIMACS funded events provided valuable opportunity for furthering our collaboration that would otherwise have been impossible."

Contributions to Other Disciplines

Since the "discipline" is inherently multidisciplinary, there is no separate entry in this section.

Contributions Beyond Science And Engineering

Several of the outcomes of the special focus have immediate applicability to real world problems.

Donald Hoover from the Statistics Department at Rutgers University reports that his theoretical work focuses on statistical methods for testing interventions (such as vaccines) to prevent rare outcomes such as the types that could be of concern for bioterrorism. He has continued and expanded work begun during the earlier years of this proposal on understanding development of maternal and infant HIV resistance to antriretroviral medications that are given to prevent maternal to infant transmission of HIV in Africa. His findings, for example, conclude that girls were at higher risk of early (in utero and perinatal) HIV infection than boys and that breastfeeding by women infected with HIV was not associated with mortality or morbidity; it was associated with highly significant reductions in mortality among their children.

William DuMouchel from Lincoln Technologies spoke on "Data Mining for Drug Safety: Statistical Analyses of Spontaneous Reports and Clinical Safety Data" in the DIMACS seminar. Postmarket data on drug safety is difficult to analyze, particularly because it has no "denominator," that is, there is no record of how many times a drug was used without side effects. Data quality is also an issue, since the same event can be reported multiple times by the doctor, medical facility and the drug company itself. Successful techniques that are robust in the face of these problems need to be developed to identify potential problems with drugs already on the market.

David Madigan from the Statistics Department at Rutgers and his coauthors wrote a report on "Issues in applied statistics for public health bioterrorism surveillance using multiple data streams: research needs." The objective of this report is to provide a basis to inform decisions about priorities for developing statistical research initiatives in the field of public health surveillance for emerging threats. It was written to help public health surveillance authorities to make better judgments on building practical systems because rapid information system advances have created a vast amount of health data pouring in from various reporting sources. They describe the analytic and statistical methodologies that synchronize best with information technologies used to gather the data. Among other things, they describe a space–time statistic that has successfully been used to detect and track public health events of interest.

Several participants in this project, working through the New Jersey Universities Consortium for Homeland Security Research, were observers for the TOPOFF 3 homeland security exercise in New Jersey in April 2005. The exercise involved release of pneumonic plague with symptomatic individuals sent to hospitals and asymptomatic individuals to Points of Dispensing (PODs) to receive antibiotics. They worked with both the NJ Office of Homeland Security and Preparedness and FEMA to provide input to their After Action Reports. They wrote a paper for the Journal of Emergency Management that summarizes their observations about the exercise, with emphasis on the POD strategy.

Through this project, we have made several connections with the Centers for Disease Control and Preparedness. CDC folks have helped organize project workshops and advised us on project design. In turn, the P.I. was invited to help CDC design a mathematical modeling program. That program is now being implemented.

Contributions To Human Resources Development

Several graduate students, undergraduates, and postdoctoral researchers have participated in the program. Local graduate students and many non-local students were also involved as visitors and workshop/working group attendees.

Specifically, the following Rutgers graduate students have undertaken small research projects under support of the special focus.

- Qinhe Cheng worked on a project titled “Causaul Effects in Dropouts from General Drug Trials”
- Srinivas Maloor worked with David Madigan from the Statistics Department.
- Graduate student David Millman has worked on random graph models of how diseases spread with DIMACS visitor Michael Capalbo.

Long-term DIMACS visitor Nina Fefferman has been investigating a variety of entropy metrics using a variety of disease data sources, and temporal and spatial scales. This year she is working with the following professors, graduate students, and undergraduate students:

- Abdul-Aziz Yakubu, Mathematics, Howard University
- Devroy McFarlane, Mathematics, Howard University (graduate student)
- Ashley Crump, Mathematics, Howard University (undergraduate student)
- Asamoah Nkwanta, Mathematics, Morgan State University
- Anthony Ogbuka, Biological Sciences, Morgan State University
- Nakeya Williams, Mathematics, Morgan State University

This project involves teams of faculty and students from two minority-serving institutions, Howard University and Morgan State University, working at DIMACS for the summer. They are funded through a Department of Homeland Security program that aims at getting faculty and students at MSIs to work on homeland security research problems. The project was developed as a result of our NSF project and is closely tied to it.

This project hosted and trained a postdoctoral fellow from National University of Singapore, Kah Loon Ng. Regarding Ng, Nina Fefferman (DIMACS) said “During his post-doctoral work at DIMACS, Kah Loon Ng and I began work together investigating the role of individual-level behavioral choice on the stability of social networks. Since social network contacts provide the means of transmission of infectious disease, we have been examining the divergent disease spread processes that emerge from the variations in stability. Our first paper in this subject was published in January of this year in the biology journal *Annales Zoologici Fennici.* This paper focused on the diversity of stable structures that could influence the evolution of dynamic social networks. Building on that publication, our second paper is already awaiting a final decision after having been provisionally accepted by the physics journal *Physical Reviews E*. In this paper, we demonstrated that disease processes on static networks was insufficient (in many cases) to accurately approximate infection spread on networks with shifting social contacts. As many of the leading studies in this field have relied on the assumption that findings on static networks are easily generalizable to any network sharing similar, average graph theoretic properties, we hope and expect that our work together at DIMACS will have a profound effect on the cutting edge of research in the field of network-based epidemiology. We are also in the process of preparing further papers building on these earlier ideas. Most fundamentally, we are examining whether or not differently behaving, but occasionally interacting networks can yield different patterns of inter-population disease transmission. These studies will be among the first to explicitly examine the network epidemiology of the formation of reservoir populations for zoonotic infections, a subject of great concern, especially considering the diversity of bird behaviors, potentially influencing their social susceptibility to avian influenza. We are also attempting to expand the scope of our investigations to examine the impact of disease avoidance behaviors on the social stability of a network. This work is still in the early stages.

“In addition to our direct research together, Dr. Ng and I have been invited to present some of our findings at various departmental seminars and professional meetings. This past year, we have given invited talks to the meeting of the DIMACS Working Group on “Spatio-temporal and Network Modelling of Diseases.” Based on our presentation at this meeting, we have initiated collaboration with Prof. Martina Morris, to explore the potential interface between her research and our own. Our work was also presented at an invited seminar talk to the Princeton University Dept. of Ecology and Evolutionary Biology and at an invited seminar talk at the Rutgers Univ. Dept of Ecology and Evolution.”

Maite Severins from the Department of Farm Animal Health at Utrecht University in the Netherlands said “I am a Phd student at Utrecht University and attended the DIMACS Workshop on Immuno-epidemiology on December 11 - 13, 2006. This has recently led to a collaboration between myself and Sander van Noort, PhD student at the Theoretical Epidemiology group of the Instituto Gulbenkian de CiEAncia in Portugal, who also attended the DIMACS workshop.”

Lastly, the workshops are also helping participants improve their teaching activities. For example, Gavin Welch from the Epidemiology and Biometry Core at Eastern Virginia Medical School wrote that after attending the Comp-Epi workshops he used several examples in his classes that he learned about by talking with other participants at DIMACS workshops.

Donald Hoover from the Statistics Department at Rutgers University has collaborated with Ph.D. statistics students in activities related to this grant. His theoretical work focuses on statistical methods for testing interventions (such as vaccines) to prevent rare outcomes such as the types that could be of concern for bioterrorism. This has lead to a PhD thesis completed by his student Qi Xia. An article from this thesis was accepted for publication in the *Journal for Biopharmaceutical Statistics* and follow up articles form the thesis are planned. He has begun work with a new Ph.D. student in the Department of Statistics, Jing Shi, to expand on these ideas.