Published in final edited form as:

Environ Ecol Stat. 2008 ; 15(3): 259–263. doi:10.1007/s10651-007-0058-4.

Statistics in disease ecology:

introduction to a special issue

Lance A. Waller

Department of Biostatistics, Rollins School of Public Health, Emory University, 1518 Clifton Road NE, Atlanta, GA 30322, USA e-mail: lwaller@sph.emory.edu

Abstract

The three papers included in this special issue represent a set of presentations in an invited session on disease ecology at the 2005 Spring Meeting of the Eastern North American Region of the International Biometric Society. The papers each address statistical estimation and inference for particular components of different disease processes and, taken together, illustrate the breadth of statistical issues arising in the study of the ecology and public health impact of disease. As an introduction, we provide a very brief overview of the area of “disease ecology”, a variety of synonyms addressing different aspects of disease ecology, and present a schematic structure illustrating general components of the underlying disease process, data collection issues, and different disciplinary perspectives ranging from microbiology to public health surveillance.

1 Placing the papers in context: a very brief overview of disease ecology

For our purposes, the term *disease ecology* summarizes the complex interactions between disease incidence in some host and various environmental/ecological processes influencing such incidence. One can study the disease ecology of chronic diseases with environmental drivers, but typically we focus on infectious diseases and seek descriptions of the primary route (s) of infection, the virulence of the pathogen, the role (if any) of vectors and reservoir hosts, and environmental impacts on the entire disease transmission process. We can consider the term “environmental” very broadly here, including social contact networks, transportation, and species interaction as well as climate and landcover. Such a broad definition of disease ecology necessarily includes a wide range of scientific disciplines including (but not limited to) entomology, climate change, veterinary medicine, microbiology, immunology, epidemiology, mathematical biology, and public health surveillance. Linking knowledge and models of the underlying disease process with observable data raises multiple issues in statistical inference, several of which are illustrated in the three papers within this issue.

The basic concepts of disease ecology are not new and the basic notion of studying interactions between pathogens, animals, climate, and humans appear within the scientific literature in many guises. Related concepts include Pavlovsky’s (1966) “landscape epidemiology”, i.e., the study of the impact of the physical landscape on the spread of infectious disease; Manel et al’s (2003) “landscape genetics” focusing on the role of landscape features on geographic and genetic differentiation of species into subpopulations; and the growing field of *conservation medicine* addressing the intersection of human health, animal health, and ecosystem health (Aguire et al. 2002; Weinhold 2003).

To focus discussion and provide a context for the three papers in this issue, consider the general schematic presented in Fig. 1, a modification of similar figures describing spatial epidemiology in general (Ostfeld et al. 2005) and applied to Lyme disease in particular (Waller et al. 2007). Each rectangle represents a data set describing the geo-spatial distribution of components of the disease system at a particular point in time (e.g., a “map” of locations of susceptible hosts,

or of reported cases of disease). The map of primary focus is the central map of the true locations of infected hosts of interest (e.g., human cases). This map is never directly observed, but we seek prediction of this map from a variety of perspectives and using a variety of types of data. Moving from left to right, we move from the basic ecology of the disease featuring the current geo-spatial distribution of the pathogen (where are infections now?), any vectors involved in transmission (where can transmission occur?), and the location of susceptible hosts (where could the disease spread next?). These maps are influenced individually and collectively by the local landscape and climate (e.g., some insect vectors require particular temperature and elevation ranges) represented by the double-ended arrows linking these features. Each “map” in the figure may in fact be comprised of multiple maps representing multiple host species (e.g., human and animal), multiple vector species, and multiple strains of the same pathogen. The relationships between these may be quite complex and study of the disease system requires linking population ecology and microbiology of the interacting pathogen—vector—host—environment systems across a wide range of spatiotemporal scales. Generally speaking, the ecology literature concentrates on mathematical models (deterministic or stochastic) describing these systems and their interactions in order to generate maps of predicted host cases. The accuracy of the models hinges on similarities between the locations and times of cases predicted by the model and the actual locations and times in the (unobserved) map of true host cases. In contrast, public health research typically focuses on the collection and analysis of data relating to the (again unobserved) map of true locations through the processes depicted on the right-hand side of Fig. 1. In this setting, true cases must be diagnosed and then reported to the data collection system. Both steps (diagnosis and reporting) can result in false positive and false negative case identification and are represented as “filters” in Fig. 1 so that the final data set on the right is an approximation to the true map of cases in the center of the figure. While basic, the relationships in Fig. 1 reveal the sorts of complexities involved in disease ecology and point out a spectrum of disciplinary foci ranging from ecology on the left moving toward public health surveillance on the right, with both focused on moving from endpoints toward the central, “true” map of disease incidence in the host of interest.

2 Overview of papers in the special issue

Reflective of the multidisciplinary nature of disease ecology, the three papers in this issue focus on particular aspects of the process in Fig. 1, but all acknowledge their connections to the “big picture”, and we set each in context here.

Bjørnstad and Grenfell explore the impact of spatial connectivity between communities on the timing of historical (prevaccination) measles outbreaks in the United Kingdom. The choice of measles provides a disease system with a well-known transmission system having strong temporal dynamics based on direct host-to-host contact. Their paper expands established probability transmission models to provide theoretical functional associations and direct empirical estimation of the role of spatial connectedness on a disease transmission, thereby providing novel insight into incidence patterns across an interconnected set of cities of various sizes. The conclusions reach beyond that of measles and provide tools for extension to more complex disease processes. While their focus is on models of the interactions on the left-hand (ecology) side of Fig. 1, the authors recognize the role of the data “filters” on the right-hand (public health) side and incorporate compensation within the model.

Walsh and colleagues focus on statistical estimation issues addressing the double-ended connection between vector abundance and climate variables on the left-hand side of Fig. 1. Of particular importance is the use of novel approaches for variable selection and lag identification linking a long list of temporally specific climate measures to the ecological samples of vector abundance. Their work also illustrates that disease ecology often involves multiple, parallel versions of the concepts illustrated in Fig. 1, most notably that different vector species for the

same family of diseases (here arboviruses), can have very different specific connections between the conceptual categories illustrated in Fig. 1, for example, different associations based different overwintering behavioral patterns between the two vector species.

Completing the special issue, Johnson provides a detailed description of public health surveillance of West Nile Virus in New York State, using concepts and knowledge of the multiple disease hosts and interactions on the left-hand (ecology) side of the figure to improve proactive surveillance of outbreaks on the right-hand side. Johnson provides a thorough literature review of relevant issues drawing from multiple disciplines and proposes additional analytic methods to fill existing gaps and provide further links between available data sets. Like Walsh and colleagues, Johnson's work also involves parallel elements in the process, but focuses on multiple hosts (humans and crows) and multiple reportable outcomes. The consideration of multiple hosts for West Nile Virus described in Johnson and the references therein is somewhat unique among diseases in that prediction of the human host case map does not depend on human reports, but on reports of a syndrome related to infection (death) in a parallel and more highly infected host species (crows). The accuracy of such a method depends critically on the sensitivity and specificity of underlying ecological knowledge relating to the disease of interest, again illustrating the importance and value of incorporating aspects from across the spectrum of topics illustrated in Fig. 1.

3 Discussion

Individually and collectively, this set of papers pushes methodologic frontiers in different disciplines. Bjørnstad and Grenfell bring statistical estimation of the impact of spatial connectivity to detailed stochastic models of spatiotemporal disease dynamics, Walsh and colleagues contribute insight to variable and lag selection for temporally-specific vector abundance data, and Johnson provides both thorough retrospective review and prospective proposals for statistical methods in West Nile surveillance to expand use of multiple outcome data sets across host species and syndromic outcomes. Taken together with concepts illustrated in Fig. 1, the papers provide an introduction to and illustration of the needs and directions for future statistical research in disease ecology.

Author Biography

Lance A. Waller is Professor in the Department of Biostatistics in the Rollins School of Public Health at Emory University and a Fellow of the American Statistical Association. Dr. Waller obtained his Ph.D. in Operations Research from Cornell University in 1991 with a minor in Epidemiology through the Veterinary School. His research interests involve the development and application of spatial statistical methodology to research problems in public health and ecology.

References

- Aguire, AA.; Ostfeld, RS.; Tabor, GM.; House, C.; Pearl, MC. Conservation medicine: ecological health in practice. Oxford University Press; Oxford: 2002.
- Manel S, Schwartz M, Luikart G, Taberlet P. Landscape genetics: combining landscape ecology and population genetics. *Trends Ecol Evol* 2003;18(4):189–197.
- Ostfeld RS, Glass GE, Keesing F. Spatial epidemiology: an emerging (or re-emerging) discipline. *Trends Ecol Evol* 2005;20:328–336. [PubMed: 16701389]
- Pavlovsky, EN. Natural nidality of transmissible diseases: with special reference to the landscape epidemiology of Zooanthroponoses. University of Illinois Press; 1966.

- Waller LA, Goodwin BJ, Wilson ML, Ostfeld RS, Marshall S, Hayes EB. Spatiotemporal patterns in County-level incidence and reporting of Lyme Disease in the Northeastern United States, 1990–2000. *Environ Ecol Stat* 2007;14:83–100.
- Weinhold B. Conservation medicine: combining the best of all worlds. *Environ Health Persp* 2003;111:A525–A529.

Fig. 1. Schematic highlighting several components in disease ecology bridging underlying ecological processes and reported public health data sets