Insights from disease ecology: focus on hand, foot and mouth disease in China

Jeffrey D. Stanaway

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2013

Reading Committee:
Jonathan D. Mayer, chair
Miles Logsdon
Anna Wald

Program Authorized to Offer Degree:
Public Health – Epidemiology

University of Washington

Abstract

Insights from disease ecology: focus on hand, foot and mouth disease in China

Jeffrey D. Stanaway

Chair of the Supervisory Committee:
Professor Jonathan D. Mayer
Epidemiology

Outbreaks of hand, foot and mouth disease (HFMD) have become increasingly regular in the Asia-Pacific region, and China has experienced annual epidemics each year since 2007. This project studied large-scale environmental drivers of infectious disease, with a focus on understanding HFMD epidemics in China between 2008 and 2011.

First, I assessed the potential of using associations with landscape pattern to discriminate diseases having a wild animal reservoir (*wild-zoonoses*) from those that do not, and tested the hypothesis that landscape, as measured by select land cover pattern metrics, is more strongly associated with the incidence of wild-zoonoses than with the incidence of diseases lacking a wild animal reservoir. Quasi-Poisson regression models were used to estimate county-level associations between land cover pattern metrics and the incidence of three wild-zoonoses and eight diseases lacking a wild animal reservoir, for each county in the contiguous United States. The absolute strengths of the associations between each pattern metric and each disease were compared to determine whether the strongest associations were observed with wild-zoonoses. When sorted by absolute strength of association, wild-zoonoses had the strongest associations with six of the ten pattern metrics (p=0.008), and

with all four of the metrics that measure land cover shape (p=0.002), suggesting that associations with land cover pattern metrics may offer insight into the existence of a wild animal reservoir for an emerging or otherwise poorly understood infectious agent

Second, I sought to fill important gaps in our understanding of the ecology of HFMD by looking at land cover and land cover pattern, and their associations with HFMD incidence in China. Univariate and multivariate associations were estimated between each of twelve landscape variables, plus population density, and HFMD incidence using quasi-Poisson regression models. Decreased elevation and vegetation density were significantly associated with increased rates of HFMD; and increased division, disaggregation, and diversity of land cover types were associated with increased rates of HFMD. The results suggest connections between landscape and HFMD incidence that warrant further investigation, and support previous studies that have found local transmission to be more important than distant transmission.

Finally, several studies have found associations between weather and HFMD, suggesting that climate change could have a role in the recent growth of HFMD in China. I sought to determine if climate change could underlie the recent emergence and growth of HFMD in China by developing a weather-based predictive model of HFMD and applying that model to historical climate data. When monthly climate-based HFMD predictions were regressed against calendar time, I found evidence of a significant increasing secular trend, with predicted rates for 2011 being 94% higher than those for 1982 (Incidence rate ratio (IRR): 1.937; 95% confidence interval (CI): 1.933, 1.940). Most of the increase in the predicted HFMD incidence occurred between 2002 and 2011, with predicted rates for 2011 being 49% higher than those for 2001 (IRR = 1.490; 95% CI: 1.488,1.493). Our climate-based

retrospective predictions suggest that changing climate should have made weather increasingly favorable to HFMD during our thirty-year study period and we find that the data are compatible with climate change playing a role in the recent growth of HFMD in China.

CONTENTS

List of Figures

List of Tables

List of Equations

List of Abbreviations

°F	Degrees Fahrenheit
AFP	Acute flaccid paralysis
AIC	Akaike information criterion
AIC-Δ	Decrease in AIC
AR 1	Autoregressive, first order
CA16	Coxsakievirus A16
CCDC	Chinese Center for Disease Control and Prevention
CDC	United States Centers for Disease Control and Prevention
CF	Case fatality
CHLYM	Chlamydia trachomatis
CI	Confidence interval
CIRCLE	Related circumscribing circle
CONTAG	Contagion
CRYPT	Cryptosporidiosis

DEM	Digital Elevation Model
ECOLI	E. Coli O157:H7
EID	Emerging infectious disease
EV71	Enterovirus 71
FIPS	Federal Information Processing Standard
FRAC	Fractal index
GB	Guo Biao
GEE	Generalized estimating equations
GIMMS	Global Inventory Modeling and Mapping Studies
GONN	Gonorrhea
GSOD	Global summary of the day
GYRATE	Radius of gyration
HFMD	Hand, foot and mouth disease
IQR	Interquartile range
IR	Incidence rate
IRR	Incidence rate ratio

km	kilometer
LEGIO	Legionellosis
LSI	Landscape shape index
LYME	Lyme disease
m	Meters
MAE	Mean absolute error
MAUP	Modifiable areal unit problem
MR	Mortality rate
NDVI	Normalized Difference Vegetation Index
NIRR	Normalized incidence rate ratio
NOAA	National Oceanic and Atmospheric Administration
PARA	Perimeter-to-area ratio
PD	Patch density
PLADJ	Percentage of like adjacencies
RMSF	Rocky Mountain Spotted Fever
SD	Standard deviation

SHAPE	Shape index
SHIG	Shigellosis
SIDI	Simpson's diversity index
SRTM	Shuttle Radar Topography Mission
STRA	Invasive group A Streptococcal disease
SYPH	Syphilis
UN LCCS	United Nations Land Cover Classification System
USGS	United States Geological Survey
WNILE	West Nile Virus

Acknowledgements

My committee members have been a constant source of knowledge and guidance, and have responded to even my most eccentric ideas with less snickering and fewer raised brows than I imagined possible. Thanks to Jonathan Mayer for taking me on as an advisee, five-years of guidance, weekend meetings, and asking the questions that have sparked some of my best ideas; Miles Logsdon for stepping outside of his field, GIS and remote sensing guruship, and our always fun conversations; Steve Self for involving me with the HFMD study group and helping me navigate data access issues; Jon Wakefield for having so many needed answers and (nearly) always countering my tendency for needless complexity; and Anna Wald for hiring me sight-unseen, mentorship, great advice, and supernatural editing skills.

Beyond my committee, several people at UW and Fred Hutch have offered invaluable help along the way: Lena Yao, Yang Yang, Xi "Cici" Chen, Ying Zhang and Lily Zhang have organized meetings, managed data, pointed me in the right direction, helped out with finicky issues in R, and otherwise made my life easier.

Of course, a dissertation with no data is...well, not much of a dissertation. And so, I must recognize the generous people at both the Chinese Center for Disease Control and Prevention (CCDC) and the US Centers for Disease Control and Prevention (CDC) who provided the surveillance data for my dissertation. Thanks to Jennifer Lehman, Greg Pierce and Michael Wodajo at the CDC for providing the National Notifiable Disease data that I needed for the validation study. Thanks to Dr. Yu Wang, director general of China CDC, and Dr. Zijian Feng, Director of the Office of Disease Control and Emergency Response, for allowing me to use the HFMD surveillance data, and for their work in

developing and maintaining the research collaboration between China CDC and Fred Hutch/UW.

Finally, thanks to those who have offered me the research and teaching assistantships that have financially supported me throughout my doctoral studies: Anna Wald, Victoria Holt, Tom Koepsell, Noel Weiss, Chris Murray, Steve Hawes, Beth Mueller, and Norm Breslow.

Dedication

For Elizabeth & Naomi

SECTION A: INTRODUCTION

I. BACKGROUND

Hand, foot and mouth disease (HFMD) is a syndrome that occurs in a small proportion of people infected with non-polio enteroviruses. It is characterized by flu-like symptoms, rash on the hands and feet, oral lesions, loss of appetite, vomiting and diarrhea. Severe cases can produce neurological complications, including acute flaccid paralysis (AFP), cardiopulmonary complications and death. HFMD is most commonly observed among children; and enterovirus 71 (EV71) and coxsakievirus A16 (CA16) are the most commonly implicated pathogens (heretofore referred to as *HFMD agents*). Transmission is thought to occur through fecal-oral routes and via respiratory droplets.[1]

Since the mid-1990's HFMD epidemics have become regular in the Asia-Pacific region and may be increasing in size and severity. Large-scale outbreaks of neurologically complicated HFMD have been reported in Malaysia (in 1997, 2000, 2003 and 2005), Taiwan (1998, 2000 and 2001), Australia (1999 and 2000), Singapore (2000 and 2006), Brunei (2006) and China (2007 onward).[1] Since 2007, China has experienced large-scale annual epidemics, prompting the Chinese Ministry of Health to add HFMD to its list of mandated notifiable diseases in 2008.[2,3]

The Chinese Center for Disease Control and Prevention (CCDC) began systematic surveillance of HFMD in 2008 and, in 2009 they initiated collaboration with the University of Washington and the Fred Hutchinson Cancer Research Center. I became involved with the Seattle-based team in November of 2009, contributing to work on two grant proposals and a study modeling the spread of HFMD outbreaks in China.[4] This study found, among

other things, strong associations between meteorological variables and HFMD incidence in China. It also noted an annual northward movement of epidemics starting near the end of winter, with retraction to southern latitudes in autumn. Thinking about these results within the larger context of our understanding of HFMD—and the viral infections that cause the disease—led me to questions about the underlying ecology of the disease: most notably, I was curious about the potential role of climate change in the recent growth of HFMD in the Asia-Pacific region, and about the underlying drivers of the observed seasonality of HFMD outbreaks. This dissertation is the product of those questions and it explores land cover pattern and climate as drivers of infectious disease incidence over space and time, with a focus on HFMD in China.

II. Rationale

Climate change and HFMD

Several studies have found associations between meteorological variables and HFMD incidence, including positive associations between temperature and HFMD—though there may be a threshold temperature above which HFMD declines.[5-15] These observed associations between weather and HFMD suggest that climate change could potentially underlie the recently observed growth of HFMD in the Asia-Pacific region and the newly established seasonal pattern of HFMD epidemics in China. While Ma et al[6] and Urashima et al[11] both considered the potential for a connection between climate change and HFMD, I have found no study that has directly investigated what role climate change may have played. With that, I

hoped to retrospectively investigate what role climatic changes may have had in the recent growth of HFMD in China.

HFMD's curious seasonality

HFMD exhibits dramatic seasonality, almost certainly due to its association with weather, with the most notable peaks in incidence occurring near summer.[5-7,16-18] Seasonal variability is most dramatic at higher latitudes and, while the overall pattern of increased incidence in the summer remains, the degree of seasonal variability is less dramatic at lower latitudes.[5,16,19] This pattern is consistent with the reported associations between meteorological variables and HFMD incidence. That HFMD exhibits pronounced seasonality is neither unique nor terribly surprising, in itself, since we observe pronounced seasonality with many infections. Still, considering our understanding of HFMD's epidemiology, I found the observed pattern curious and difficult to explain. It is well accepted that HFMD agents (most notably EV71 and CA16) are transmitted directly through both fecal-oral and oral-oral routes; with evidence suggesting that water-borne, food-borne and indirect transmission through fomites are also possible.[1,5,6,19,20] If we believe that HFMD agents may be transmitted via *only* these routes, and between *only* humans, then we must classify HFMD as a non-vectored anthroponotic disease. And this, I felt, created a problem: how can we explain profound and consistent summertime peaks in the incidence of a viral agent within this transmission paradigm?

I have found few authors who have suggested a potential mechanism for HFMD's seasonality and I believe those proposed mechanisms conflict with evidence and our

understanding of the disease. Some have suggested that summer weather may improve the viability and survivability HFMD agents in the environment[21]; however, these viruses are known to survive longer with cool temperatures than with warmer ones. Others have suggested that the summertime peak may correspond with social gathering, children playing together outdoors, or some other driver of increased mixing between infected and susceptible hosts.[6] And evidence does suggest that crowding and social gathering increase HFMD transmission.[22] Still, playing outdoors would seem to offer less potential for transmission than would playing indoors, as children do in wintertime. Also, few social gathering events have a shared timing across all areas affected by HFMD. Focusing on children, among whom HFMD is the greatest problem, the most notable such event is school; and while there is evidence that bringing children together for school does increase HFMD risk,[5,7] the infection's peak season coincides with a period when children are typically out of school and on summer break. I therefore find the case for these mechanisms unpersuasive. Some spatial and temporal variability is likely the result of variability in public health practices and infrastructures. These issues certainly affect the disease's spatial distribution and likely drive secular temporal trends but, as these are not seasonal drivers, I don't see how they can explain HFMD's seasonality.

From the perspective of established frameworks of infectious disease seasonality, therefore, it seems difficult to explain the pattern observed with HFMD: seasonal behaviors, pathogen survival, and seasonal changes in host immunity are all far more consistent with increased transmission in the winter than in the summer. I, consequently, came to believe that either HFMD's seasonality is driven by some novel and as of yet unidentified

mechanism, or that our understanding of its epidemiology is incomplete. And, with that, I suggested that one possible explanation is the existence of an unknown animal reservoir. Under this scenario, the northward movement of HFMD with the end of winter, and its southward retreat in the autumn, could result from the corresponding migratory movement of the animal reservoir; and associations with other meteorological variables could result from the effects of weather on the reservoir population (*e.g.* wet years could increase HFMD risk by increasing the reservoir's food supply and, thereby, amplifying the reservoir population). Though clearly not the only plausible explanation, the existence of an animal reservoir is plausible. And, while no direct evidence exists for an animal reservoir, other species of enteroviruses have been isolated from a variety of animals, suggesting the biological plausibility of inter-species transmission of enteroviruses.[23-26] Moreover, with 60 to 75% of emerging infectious diseases being zoonoses,[27-29] the existence of an unknown animal reservoir for HFMD agents seemed possible, even if improbable.

Definitively establishing the existence of an animal reservoir for HFMD agents would require data supporting an epidemiologic connection between human infections and the proposed reservoir species, and replicated laboratory confirmation of the relevant enterovirus (*e.g.* EV71 or CA16) infection occurring in the proposed reservoir species. This approach to reservoir discovery would be long and expensive—it is similar to approach used in the roughly three-decade-long hunt for the natural reservoir of Ebola[30,31]—and the inductive rationale for an animal reservoir was too weak to warrant such a study in the case of HFMD. First, I felt that studies of existing data were needed and, if the results of these studies strongly suggested the existence of an animal reservoir, then large-scale field studies might be

warranted. With that, I thought that the rationale was adequate to warrant an investigation that could shed light onto the issue through an analysis of existing data. I, consequently, began thinking about how existing data could be used to tackle this question.

Applying existing data to reservoir discovery

Conventionally, we can classify infectious agents by their means of transmission into those that are transmitted directly between two hosts (*e.g.* fecal-oral, sexual, or airborne transmission) and those that are transmitted indirectly (*e.g.* vector-borne, water-borne or food-borne transmission). Some agents may be transmitted either directly or indirectly (*e.g.* norovirus, which may be transmitted directly, through aerosolized feces or vomit, or indirectly, through contaminated food). Transmission can then be subdivided between *anthroponotic*, between humans, and *zoonotic*, between humans and other animals. This, then, gives us four categories: *directly transmitted anthroponoses*, *directly transmitted zoonoses*, *indirectly transmitted anthroponoses*, and *indirectly transmitted zoonoses*. From the perspective of a component-based transmission model, we can describe these four transmission categories as combinations of four components: agent, host, animal reservoir, and indirect vehicles (*e.g.* vectors and fomites). Agent and host compose the minimum set of necessary components, and are the requisite components for direct anthroponotic transmission. We can also see that indirect transmission, which includes some transmission vehicle (*e.g.* vector or fomite), requires a more complex model than does direct transmission; and zoonoses, which include an animal reservoir, require a more complex model then do anthroponoses.[32-34] Within this framework, the distribution of a disease arises from the joint distribution of both the

components of its transmission model and the factors influencing the interactions between these components. Simply put, diseases will occur only where and when their requisite transmission components exist and interact; and, beyond the presence of all the requisite components and component-interactions, incidence will be greatest in those places and times where components are most prevalent and their interactions are least restricted. With that, transmission requires host-host interactions—in the case of directly transmitted agents—and reservoir-host interactions—in the case of indirectly transmitted agents. *Contact rate* describes the frequency of such interactions and, all other things being equal, higher contact rates generate higher transmission rates.[33,34] Thus, any factor that influences the contact rate between humans and animal reservoirs, but not human-to-human contact rates, should be more strongly associated with the incidence of zoonoses than with the incidence of anthroponoses.

I hypothesized that patterns of land cover should strongly influence contact rates between infectious wild-animal reservoirs and susceptible humans, but should not strongly influence contact rates between infectious and susceptible humans. Most simply, susceptible humans and infective reservoir animals must interact for zoonotic transmission to occur [35,36]; and the potential for that interaction is at least partially governed by landscape. Similarly, Lloyd-Smith et al [37] described the rate of zoonotic transmission as being governed by the rate, duration, and proximity of contact between the reservoir and humans, among other things. Because landscape influences the potential for such contact, landscape should, in turn, also be associated with the incidence of zoonotic diseases. Empirically, this assumption is borne-out by evidence of increased mixing between humans and deer ticks resulting in

higher rates of Lyme disease.[29,35] Likewise, factors including forest fragmentation and agricultural encroachment are associated with increased risk of emerging zoonoses: for example, the emergence of Nipah virus has been linked to agricultural expansion in Malaysia; roads constructed in uninhabited areas (*e.g.* logging and mining roads) have given bushmeat hunters access to previously inaccessible wilderness and the wildlife living therein, and have thus been implicated in the first emergence of human immunodeficiency virus (HIV) and simian foamy virus.[29] In contrast, such associations would be unlikely between landscape and non-zoonotic diseases.

Accordingly, we should systematically detect strong associations between land cover pattern and the incidence of wild-zoonoses (*i.e.* diseases having a wild-animal reservoir), but not detect such systematic associations between land cover pattern and the incidence of anthroponoses. Moreover, land cover pattern can be quantified from existing remote sensing data through a variety of land cover pattern metrics, making it an excellent candidate variable for discriminating zoonotic and anthroponotic diseases using only existing data.

What emerged

I had hoped, therefore, to use associations with land cover pattern to investigate the possible existence of an animal reservoir for HFMD agents. Before attempting to do so, I planned to first validate the use of land cover pattern for discriminating zoonotic and anthroponotic diseases against well-understood diseases in the United States. If associations with pattern metrics clearly and systematically differed between zoonotic and non-zoonotic diseases in my validation, then I hoped to take a first step in evaluating the possibility of an animal reservoir

for HFMD agents: I planned to compare the nature of HFMD's associations with various land cover pattern metrics to the nature of those same associations seen with the diseases in the validation study. With further work, however, I found that my planned validation was better described as a proof-of-concept study, and that more validation and development would be needed before applying this proposed method to HFMD in China. Consequently, what began as two tightly connected studies became two separate but related studies: first, a proof-of-concept study testing the hypothesis that land cover pattern is more strongly associated with the incidence of wild-zoonoses than with the incidence of diseases that lack such a reservoir; and, second, a descriptive study looking at associations between land cover, land cover pattern and HFMD incidence in China.

III. Specific Aims

1. To explore the utility of studying associations between landscape pattern and disease incidence in determining if a disease has a wild animal reservoir; specifically, to test the hypothesis that land cover pattern metrics are more strongly associated with the incidence of wild-zoonoses than with the incidence of diseases that lack such a reservoir.

2. To better understand the ecology of HFMD by describing its associations with land cover type and land cover pattern in mainland China.

3. To determine if climate change has produced weather that is increasingly favorable to HFMD and may, thereby, underlie the recent increase in HFMD activity in mainland China.

I addressed each of these aims in a separate study. In the next section, I present each of the three studies' corresponding manuscript.

SECTION B: STUDY MANUSCRIPTS

I. LAND COVER PATTERN AND WILD-ZOONOSES

Abstract

We assessed the potential of using associations with landscape pattern to discriminate diseases having a wild animal reservoir (*wild-zoonoses*) from those that do not, and tested the hypothesis that landscape, as measured by select land cover pattern metrics, is more strongly associated with the incidence of wild-zoonoses than with the incidence of diseases lacking a wild animal reservoir. We obtained county-level incidence data from CDC National Notifiable Disease records data for three wild-zoonoses (Lyme Disease, Rocky Mountain Spotted Fever, and West Nile Virus) and eight diseases lacking a wild animal reservoir (Chlamydia trachomatis, Cryptosporidiosis, E. coli O157:H7, Gonorrhea, Legionellosis, Shigellosis, invasive group A Streptococcal disease, and Syphilis). Using GlobCover 2009 land cover data, we calculated 10 land cover pattern metrics for each county in the contiguous United States and modeled their associations with the county-level incidence of each of the eleven diseases using quasi-Poisson regression. Finally, we compared the absolute strengths of the associations between each pattern metric and each disease to determine whether the strongest associations were observed with wild-zoonoses. When sorted by absolute strength of association, wild-zoonoses had the strongest associations with six of the ten pattern metrics (p=0.008), and with all four of the metrics that measure land cover shape (p=0.002). While more work is needed to refine this method, our results suggest that associations with land cover pattern metrics may offer insight into the existence of a wild animal reservoir for an emerging or otherwise poorly understood infectious agent.

Background

Sixty to 75% of all emerging infectious diseases (EID) are estimated to be zoonotic,[27-29] and undetected animal reservoirs may exist for established infectious agents. Given limited resources and the high cost of large-scale wildlife surveillance, a rapid and inexpensive screening tool may expedite the discovery of animal reservoirs, inform research and surveillance priorities, and allow the public health community to respond to EIDs more effectively. To address this need, we propose a novel method through which surveillance data and remote sensing imagery may be used to investigate the existence of a wild animal reservoir for any poorly understood infectious agent, and conduct an exploratory analysis to ascertain the validity of such a method.

We assume that mixing between humans and wild animals increases zoonotic disease transmission, and that mixing is governed—at least partially—by landscape. Most simply, proximity between humans, infectious agents, and natural reservoirs is necessary for zoonotic transmission to occur.[35,36] Similarly, Lloyd-Smith et al[37] described the rate of zoonotic transmission as being governed by the rate, duration, and proximity of contact between the reservoir and humans, among other things. Because landscape influences the potential for such contact, landscape should, in turn, also be associated with the incidence of zoonotic diseases. Empirically, this assumption is borne-out by the historical evidence of increased mixing between humans and deer ticks resulting in higher rates of Lyme disease.[29,35] Likewise, factors including forest fragmentation and agricultural encroachment are associated with increased risk of emerging zoonoses.[29] In contrast, such associations would be unlikely between landscape and non-zoonotic diseases.

A variety of land cover pattern metrics have been developed to quantify some aspect of land cover pattern: Simpson's diversity index, for example, quantifies the diversity of land cover types seen in a given landscape. Within the context of pattern metrics, a patch is a contiguous area of a single land cover type (*e.g.* a contiguous urban patch), a class is a category of land cover type (*e.g.* deciduous forest, urban, or open water), and a landscape is the total land area included in an analysis and comprises multiple patches and classes.[5,38]

In this study we test the hypothesis that landscape, as measured by a variety of land cover pattern metrics, is more strongly associated with the incidence of wild-zoonoses (*i.e.* diseases having a wild animal reservoir) than with the incidence of diseases lacking a wild animal reservoir. We then discuss the utility of different pattern metrics in predicting whether or not a disease has a wild animal reservoir. While we expect this method will imperfectly discriminate wild-zoonoses, it does offer the potential to provide a relatively fast and inexpensive means of screening suspected zoonotic agents for further study. The need for quick and efficient responses to emerging infections[39] makes this an attractive and promising approach.

Methods

Land Cover Data Sources & Processing

We obtained data on land cover from the GlobCover 2009 land cover map.[40] This is a 300 meter (m) resolution map in which land cover is categorized into 22 classes as defined by the United Nations Land Cover Classification System (UN LCCS). Of these, one class corresponds to artificial surfaces and urban areas, three classes correspond to areas that are either entirely or primarily agricultural, and the remaining 18 classes correspond to areas that

Table 1.1: Land cover classes and their assigned values under the two-class scheme

Value	GlobCover class definition [50]	Two-classes
11	Post-flooding or irrigated croplands	Developed
14	Rainfed croplands	Developed
20	Mosaic Cropland (50-70%) / Vegetation (grassland, shrubland, forest) (20-50%)	Developed
30	Mosaic Vegetation (grassland, shrubland, forest) (50-70%) / Cropland (20-50%)	Undeveloped
40	Closed to open (>15%) broadleaved evergreen and/or semi-deciduous forest (>5m)	Undeveloped
50	Closed (>40%) broadleaved deciduous forest (>5m)	Undeveloped
60	Open (15-40%) broadleaved deciduous forest (>5m)	Undeveloped
70	Closed (>40%) needleleaved evergreen forest (>5m)	Undeveloped
90	Open (15-40%) needleleaved deciduous or evergreen forest (>5m)	Undeveloped
100	Closed to open (>15%) mixed broadleaved and needleleaved forest (>5m)	Undeveloped
110	Mosaic Forest/Shrubland (50-70%) / Grassland (20-50%)	Undeveloped
120	Mosaic Grassland (50-70%) / Forest/Shrubland (20-50%)	Undeveloped
130	Closed to open (>15%) shrubland (<5m)	Undeveloped
140	Closed to open (>15%) grassland	Undeveloped
150	Sparse (>15%) vegetation (woody vegetation, shrubs, grassland)	Undeveloped
160	Closed (>40%) broadleaved forest regularly flooded - Fresh water	Undeveloped
170	Closed (>40%) broadleaved semi-deciduous and/or evergreen forest regularly flooded - Saline water	Undeveloped
180	Closed to open (>15%) vegetation (grassland, shrubland, woody vegetation) on regularly flooded or waterlogged soil - Fresh, brackish or saline water	Undeveloped
190	Artificial surfaces and associated areas (urban areas >50%)	Developed
200	Bare areas	Undeveloped
210	Water bodies	Undeveloped
220	Permanent snow and ice	Undeveloped

are either entirely or primarily undeveloped (Table 1.1). As edges between two undeveloped classes were not hypothesized to have any effect on the incidence of zoonotic diseases, we dichotomized the 22 native land cover classes, into only two: "natural" classes and "developed" classes. Developed classes included those comprising primarily urban, artificial and agricultural land covers (Table 1.1).

Separate county-level land cover classification raster files were extracted using the boundaries defined in national county-level cartographic boundary shapefiles from the US Census Bureau's TIGER geographic database. These county raster files were then analyzed using Fragstats 4.1[41] to calculate ten land cover pattern metrics: of these, four describe some element of patch shape (shape metrics), four describe the degree to which pixels of the same land cover class are aggregated or disaggregated (aggregation metrics), and the remaining two metrics describe patch area and diversity, respectively (Table 1.2). These metrics were selected based on their relevance to the hypothesis; and an effort was made to select diverse metrics that describe different aspects of pattern and avoid excessive redundancy. Several of the metrics chosen operate inherently at the patch-level rather than the landscape-level; these metrics can describe pattern at the landscape-level, however, through statistics summarizing their distribution across all patches in a landscape. For example, the perimeter-to-area ratio is calculated for each patch as the perimeter of that patch, divided by its area. To evaluate this characteristic at the landscape-level, though, we can use the mean value of the distribution of patch-level perimeter-to-area ratios in that landscape. The mean values of inherently patch-level metrics—including, perimeter-to-area ratio, fractal index, radius of

Table 1.2: Summary of pattern metrics, including the theoretical range of values that each metric can take, and brief descriptions.

Metric (Abbreviation)	Range	Description
Aggregation metrics		
Contagion (CONTAG)	$0 < CONTAG \leq 100$	Quantifies aggregation based on the degree to which adjacencies in a landscape are between pixels of the same or differing classes. Low values indicate that pixels of the same class are clustered together; higher values indicate that pixels of the same class are dispersed throughout the landscape and interspersed with pixels of different classes.
Landscape shape index (LSI)	$LSI \geq 1$	LSI is calculated as a quarter of the total edge in a landscape (m) divided by the square root of the area of that landscape. Higher values indicate greater disaggregation within the landscape and may be indicative of a greater patch density or greater patch shape complexity.
Patch density (PD)	$PD > 0$	The number of patches per hectare of landscape.
Percentage of like adjacencies (PLADJ)	$0 \leq PLADJ \leq 100$	The percentage of pixel adjacencies in which the adjacent pixels are of the same class. A value of zero would indicate completely disaggregated landscape (i.e. a checker board-like configuration), higher values indicating greater aggregation, and a value of 100 indicating a landscape in which all pixels are of the same class.
Area metric		
Radius of gyration, mean (GYRATE)	$GYRATE \geq 0$	The mean distance between each pixel and the centroid of the patch that comprises that pixel, in meters. Larger values indicate more expansive patches.
Diversity metric		
Simpson's diversity index (SIDI)	$0 \leq SIDI < 1$	"SIDI equals 1 minus the sum, across all patch types, of the proportional abundance if each patch type squared" and it "represents the probability that any 2 pixels selected at random would be different patch types." [8]
Shape metrics		
Fractal Index, mean (FRAC)	$1 \leq FRAC \leq 2$	A modification version of PARA that quantifies shape complexity but is unaffected by patch size. For each patch It is calculated 2*ln(0.25p)/ln(a), where is p and a are the patch perimeter and area, respectively. Higher values indicate greater shape complexity.
Perimeter-area ratio, mean (PARA)	$PARA > 0$	The mean perimeter-to-area ratio (m/m^2) of all patches in a landscape. Higher values may indicate either greater shape complexity or smaller patch-size.
Related circumscribing circle, mean (CIRCLE)	$0 \leq CIRCLE < 1$	"CIRCLE equals 1 minus patch area (m^2) divided by the area (m^2) of the smallest circumscribing circle." [8] Higher values indicate patches that are slender and elongated.
Shape index, mean (SHAPE)	$SHAPE \geq 1$	A modification version of PARA that quantifies shape complexity but is unaffected by patch size. For each patch It is calculated as a quarter of the perimeter divided by the square root of the area. A value of one indicates a square patch; higher values indicate greater shape complexity.

gyration, related circumscribing circle, and shape index—were used in this study. Detailed information on these pattern metrics is available elsewhere.[5,38]

Disease Data Sources & Processing

County-level incidence data were obtained from CDC National Notifiable Disease records. These data were collected by states through state-mandated reporting, and assembled by the CDC. Diseases for this analysis were selected from the list of National Notifiable Diseases via systematic application of inclusion and exclusion criteria. Diseases were eligible for inclusion if data were available for at least five consecutive years between 2001 and 2008; if, during these five years, reporting for the disease was mandated by at least forty states; and if the national incidence during each of these years exceeded 1,000 cases. Moreover, diseases against which humans are vaccinated were also excluded. An association might exist between landscape and vaccination rates (hypothetically, vaccination rates might be higher in urban areas, for example); if this were the case, then an association between land cover pattern and disease incidence would arise through two causal paths: first, through the direct effects of landscape on transmission potential (*i.e.* the casual pathway of interest), and second, through a pathway created by the association between landscape and vaccination rates. And the presence of this second causal path, mediated by vaccination rates, would interfere with our ability to estimate the strength of any association between landscape and disease incidence that results from the causal path of interest. Moreover, lack of comprehensive national data on county-level vaccination rates precluded blocking this secondary path by statistical adjustment. The remaining diseases were then classified into three categories based on their

epidemiology within the contiguous United States[10]: those primarily transmitted to humans via a wild animal reservoir, herein referred to as *wild-zoonoses* (*n*=3); those for which transmission between humans and wild animal reservoirs occurs, but is not the primary mode of transmission, herein referred to as *secondarily zoonotic* (*n*=2); and those for which transmission between wild animals and humans is thought not to occur in the contiguous United States, herein referred to as *non-wild-zoonoses* (*n*=8). For clarity, non-wild-zoonoses comprise anthroponoses and zoonoses having a reservoir that is not a wild animal (*i.e.* zoonoses with a domestic or livestock reservoir). Given the ambiguous nature of secondarily zoonotic diseases, and the amount of misclassification that would inevitably result from classifying all cases of these diseases as either zoonotic or non-zoonotic, secondarily zoonotic diseases were excluded. All remaining diseases were included in the analysis and include three wild-zoonoses (Lyme Disease (LYME), Rocky Mountain Spotted Fever (RMSF), and West Nile Virus (WNILE)) and eight non-wild-zoonoses (Chlamydia trachomatis (CHLYM), Cryptosporidiosis (CRYPT), E. coli O157:H7 (ECOLI), Gonorrhea (GONN), Legionellosis (LEGIO), Shigellosis (SHIG), invasive group A Streptococcal disease (STRA), and Syphilis (SYPH)).

Population estimates for each county during each year of disease reporting were obtained from the US Census Bureau. For each disease and each county, the observed numbers of cases were calculated as the sum of all cases of that disease reported in that county during the five-year period. For each disease, the national incidence rate was calculated as the total number of cases reported nationally during the five-year period, divided by the sum of the annual national population estimates for each of the same five

years (*i.e.* the total number of cases, divided by the number of person-years from which those cases arose). The expected numbers of cases of each disease in each county were calculated as the product of the national incidence rate and sum of the annual population estimates for that county over the relevant five-year period. These data were merged with the county-level landscape data based on Federal Information Processing Standard (FIPS) codes.

Data Analysis

In addition to calculating national case counts and incidence rates, for each disease we calculated medians and interquartile ranges (IQR) of the distributions of both county-level case counts and county-level incidence rates. Similarly, we calculated to mean, standard deviation and range for each pattern metric. Incidence rates were mapped for each disease using county-level choropleth maps to visually assess their spatial patterns.

Separate quasi-Poisson regression models were used to estimate the associations between each pattern metric and the incidence of each disease. The models took the form of,

$$ln(\theta_i) = \alpha + \beta x_i \tag{1.1}$$

Where θ_i is the incidence rate in county i, x_i is the value of the pattern metric, and β is the log of the incidence rate ratio (IRR) associated with each one unit change in metric x. Of note, our model did not account for the spatial structure of the data. Accurate estimation of standard errors and, in turn, accurate hypothesis testing require that one account for residual spatial dependence, when it is present; however, accounting for spatial dependence can dilute the observed associations. When interest lies in the model coefficients, but no inferences are

to be made (*i.e.* when the accuracy of standard errors is less important than the accuracy of coefficients), accounting for spatial dependence may yield misleading results.[42]

For each model, the absolute value of the coefficient, β, was used to assess the absolute strength of the association between that pattern metric and disease. That is, the direction of the association was not considered in evaluating the strength of the association. The degree to which the strengths of associations between pattern metrics and disease incidence vary for zoonotic and non-zoonotic diseases was assessed separately for each metric and globally, considering associations with all metrics together.

For each metric, diseases were sorted based on the absolute strengths of their associations with that metric and assigned ranks based on their sort order: the disease with the weakest association was assigned the rank of one, and the disease with the strongest association was assigned the rank of eleven. Metrics that were more strongly associated with the three wild-zoonoses than with the eight non-zoonotic diseases (*i.e.* metrics for which the three zoonotic diseases occupied ranks nine through eleven) were considered to be perfectly consistent with our hypothesis. Thus, each metric could be viewed as a Bernoulli trial, with success defined as ranks nine through eleven being occupied by our three zoonotic diseases. If the values of the pattern metrics were independent, we could calculate the probability of seeing the observed number of successes as a simple n choose k binomial coefficient problem; however, given the lack of independence, we estimated this probability via simulation. For this, we shuffled the pattern metric variables such that the values of all metrics from a given county were randomly assigned to a different county, and then estimated the associations between each of these shuffled pattern metrics and the incidence of each disease. When all of

the associations were estimated, that iteration was complete and the counties to which the pattern metrics were assigned were again shuffled. After 1,000 iterations, the resulting distribution of number of successes was inspected. Two-sided *p*-values were estimated as the proportion of these iterations in which an outcome occurred that was at least as extreme (*i.e.* at least as far from the null hypotheses) as that observed with the real data.

Normalized incidence rate ratios (NIRR) were generated for ease of interpretation and to facilitate comparisons between metrics. For each pattern metric, *m*, and each disease, *d*, the corresponding NIRR was calculated as the coefficient of the regression model, times the standard deviation, σ, of that metric, exponentiated:

$$NIRR_{dm} = e^{\beta_{dm} \cdot \sigma_m} \tag{1.2}$$

Thus, each NIRR represents the ratio of the expected incidence rates in two counties having values of a given pattern metric that differ by one standard deviation.

ArcGIS version 10.0 (ESRI, Redlands, CA) was used for geospatial data management tasks and mapping, with scripting in Python 2.6; R version 2.12.0 was used for regression modeling and simulations; and Stata version 11.2 (StataCorp, College Station, Texas) was used for general data management and descriptive analyses.

Results

We used surveillance data from 2004 through 2008 for all diseases except E. coli O157:H7, for which we used data from 2001 through 2005. Chlamydia was the most commonly

Table 1.3: Disease summary statistics

Disease	Reporting years	Wild-Zoonosis	National Case Count†	National IR‡	Cases per County† Median (IQR)	County IRs‡ Median (IQR)
Chlamydia	2004 - 2008	No	5,150,121	347.1	260 (81, 856)	178.12 (109.4, 316.5)
Cryptosporidiosis	2004 - 2008	No	35,285	2.4	2 (0, 7)	1.05 (0, 3.2)
E. coli O157:H7	2001 - 2005	No	17,037	1.2	1 (0, 4)	0.65 (0, 2.0)
Gonorrhea	2004 - 2008	No	1,696,361	114.3	42 (9, 222)	27.32 (11.0, 91.0)
Invasive group A Streptococcal disease	2004 - 2008	No	24,217	1.6	1 (0, 4)	0.70 (0, 2.1)
Legionellosis	2004 - 2008	No	13,059	0.9	0 (0, 2)	0.00 (0, 0.8)
Lyme Disease	2004 - 2008	Yes	122,175	8.2	0 (0, 3)	0.00 (0, 1.2)
Rocky Mountain Spotted Fever	2004 - 2008	Yes	10,575	0.7	0 (0, 2)	0.00 (0, 1.0)
Shigellosis	2004 - 2008	No	87,971	5.9	2 (0, 12)	1.63 (0, 5.1)
Syphilis	2004 - 2008	No	189,477	12.8	3 (0, 14)	1.81 (0, 5.3)
West Nile Virus	2004 - 2008	Yes	14,794	1.0	0 (0, 2)	0.00 (0, 1.2)

†Cumulative number of cases over the five-year reporting period

‡ Incidence Rates per 100,000 person-years

occurring infection, with over 5 million cases reported during five-years. Lyme disease was the most commonly occurring zoonotic infection, with over 122 thousand cases reported during five years (Table 1.3). Summary statistics describing the distribution of pattern metrics are given in Table 1.4.

When sorted by absolute strength of their associations with each pattern metric, wild-zoonoses were far more likely to occupy higher ranks than were other diseases. Moreover, wild zoonoses occupied ranks nine through eleven (*i.e.* had the strongest associations) for 6 of 10 pattern metrics (Figures 1.1–1.3). Based on simulations, under the null hypothesis (*i.e.* wild-zoonoses and non-wild-zoonoses are equally likely to be most strongly associated with any given pattern metric), the probability of observing 6 or more successes out of 10 trials is approximately 0.004 (two-tailed p=0.008).

Table 1.4: Means, standard deviations and ranges of county-level pattern metrics for all counties in the contiguous United States.

Metric	Mean	(SD)	Range
Contagion (%)	72.09	(26.69)	0.00 – 99.97
Fractal index, mean	1.03	(0.01)	1.00 – 1.07
Landscape shape index	5.80	(4.67)	1.07 – 26.50
Patch density (patches/hectare)	0.08	(0.08)	0.00 – 0.82
Percentage of like adjacencies (%)	94.25	(6.20)	70.20 – 100.00
Perimeter-to-area ratio, mean (m/m^2)	92.38	(13.87)	0.37 – 120.05
Radius of gyration, mean (meters)	1083.41	(2388.54)	250.39 – 65509.54
Related circumscribing circle, mean	0.46	(0.09)	0.08 – 0.76
Shape index, mean	1.26	(0.12)	1.01 – 2.13
Simpson's diversity index	0.13	(0.15)	0.00 – 0.50

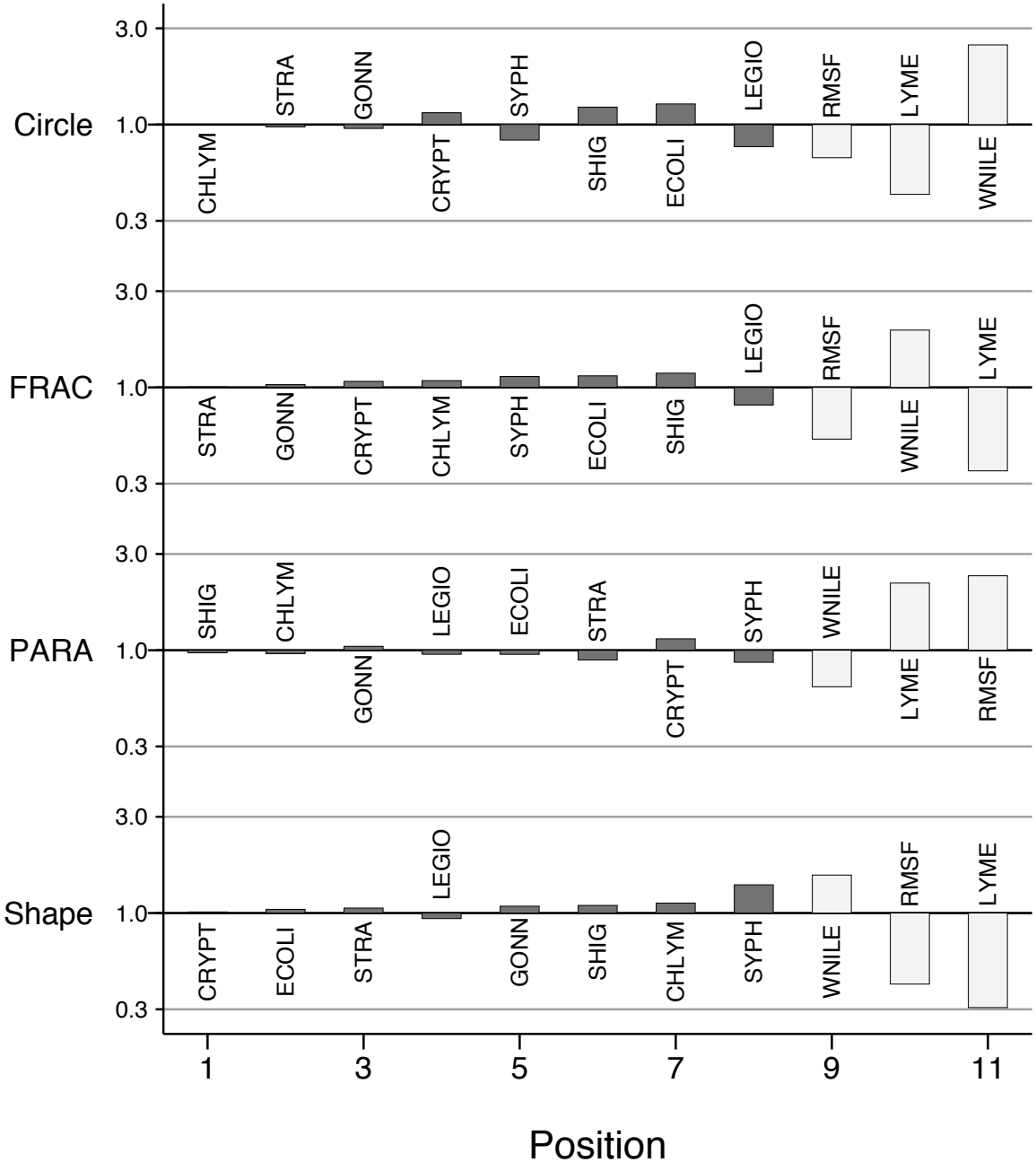

Figure 1.1: Normalized incidence rate ratios (NIRR) for associations between each shape metric and each disease. Associations with wild-zoonoses and diseases without wild-animal reservoirs are represented by light gray and dark gray bars, respectively. Note that bar heights are shown on a log scale.

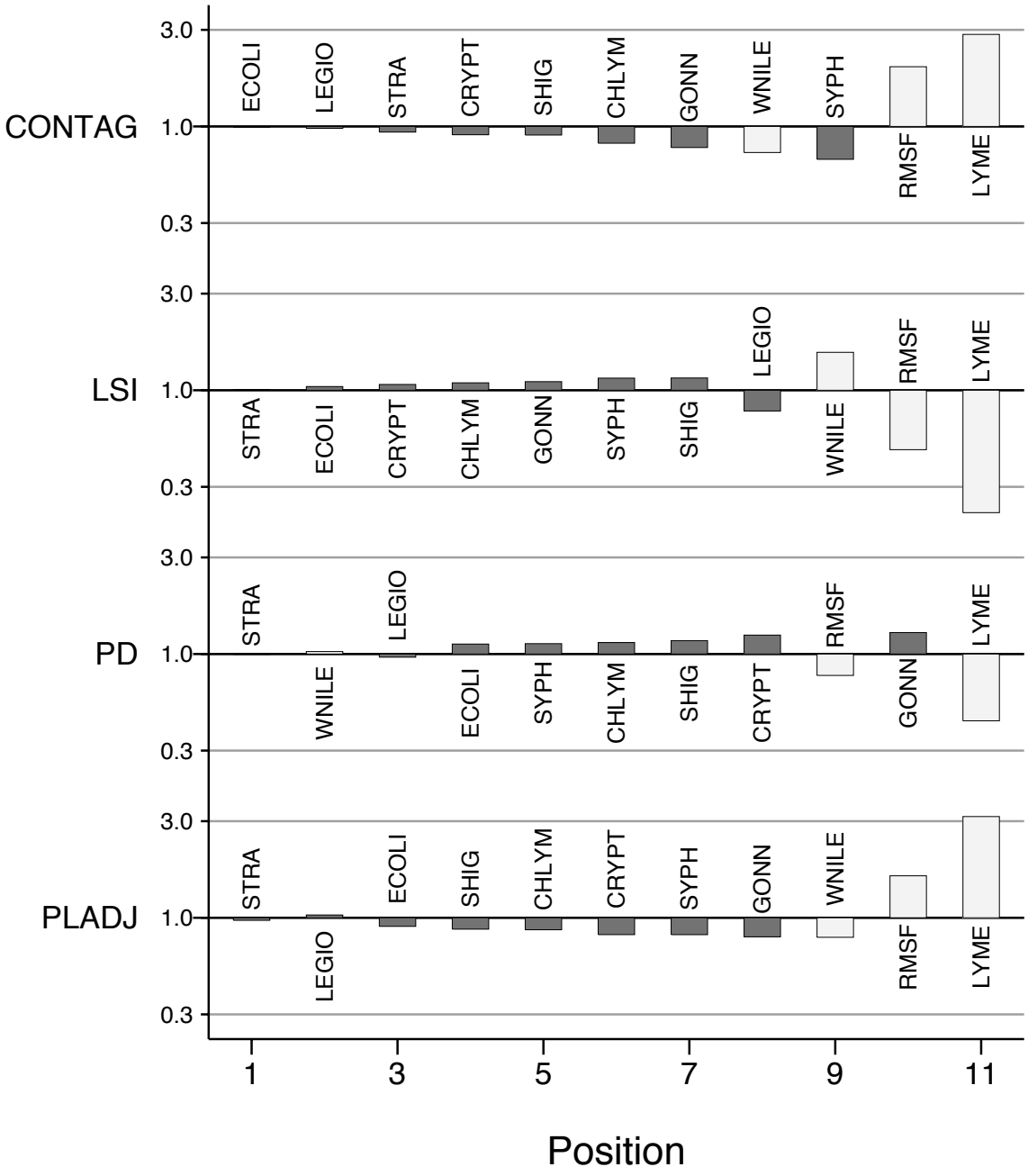

Figure 1.2: Normalized incidence rate ratios (NIRR) for associations between each aggregation metric and each disease. Associations with wild-zoonoses and diseases without wild-animal reservoirs are represented by light gray and dark gray bars, respectively. Note that bar heights are shown on a log scale.

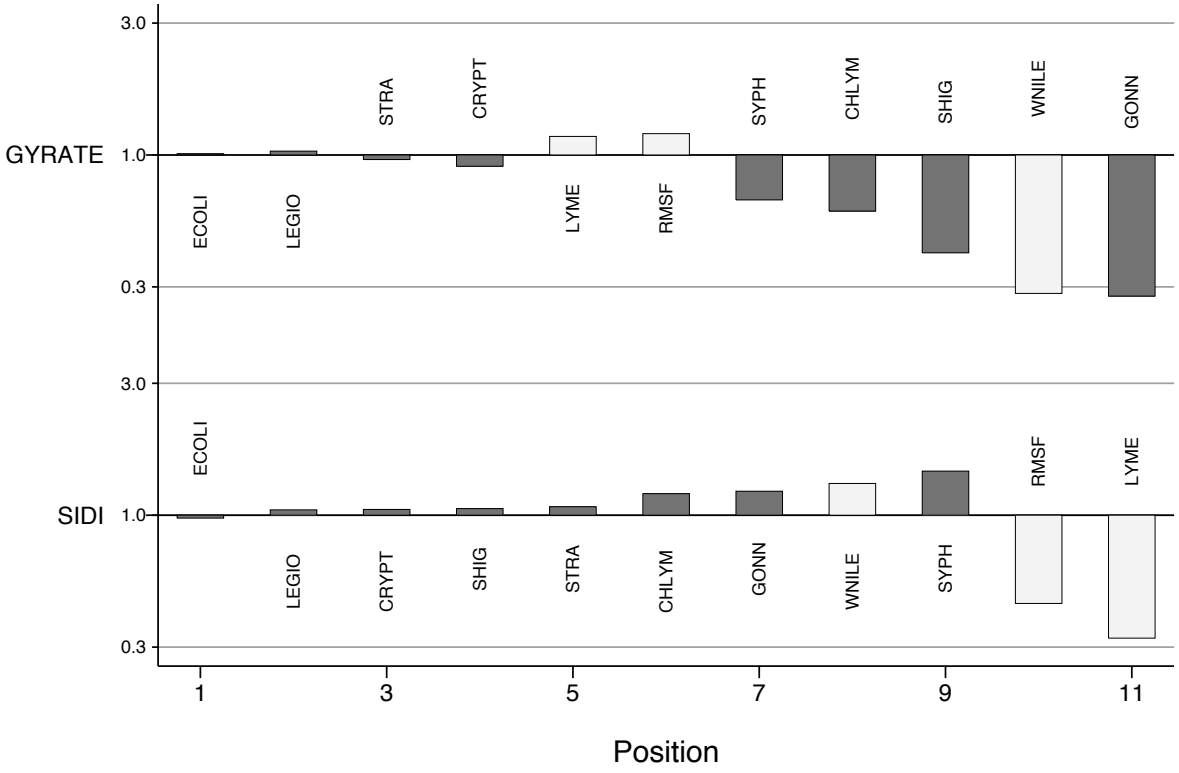

Figure 1.3: Normalized incidence rate ratios (NIRR) for associations between area and diversity metrics and each disease. Associations with wild-zoonoses and diseases without wild-animal reservoirs are represented by light gray and dark gray bars, respectively. Note that bar heights are shown on a log scale.

When the strengths of these associations were viewed by category of pattern metric, wild-zoonoses had the strongest associations with all four shape metrics (p=0.002)(Figure 1.1) and with two of the four aggregation metrics: percentage of like adjacencies and landscape shape index (p=0.08)(Figure 1.2). Conversely, the three wild-zoonoses were not the diseases that were most strongly associated with the remaining metrics (radius of gyration and Simpson's Diversity Index)(Figure 1.3).

Discussion

Our analysis found that most land cover pattern metrics were more strongly associated with rates of wild-zoonoses than they were with rates of non-zoonotic diseases, and that these differences were very unlikely to result from chance alone. When assessing the ability of a pattern metric to discriminate between zoonotic and non-zoonotic diseases based on rank order of the absolute strength of association, as a group, shape metrics outperformed those from other categories. Both of the non-shape metrics with which zoonotic diseases has the strongest associations (percentage of like adjacencies and landscape shape index) are classified as aggregation metrics, but landscape shape index does also capture patch shape complexity and, like several shape metrics, it is a function of the ratio between total length of patch edges found in a landscape to total area of that landscape. This supports the underlying rationale for using land cover pattern to identify wild-zoonoses: greater patch edge offers greater opportunity for mixing between susceptible humans and infected wild-animal reservoirs; and other aspects of land cover pattern, though potentially relevant, have a weaker influence on contact rates between humans and wild-animal reservoirs.

For each metric the direction of the associations were the same with both RMSF and LYME; and associations with WNILE were always in the opposite direction. Importantly, the similar associations seen between various pattern metrics and both RMSF and LYME do not arise from the diseases simply being collocated. Conversely, the county-level incidence rates of the two diseases are essentially uncorrelated ($\rho = -0.032$). Given the lack of colocation, the most plausible explanation for their similar associations with landscape pattern metrics is their common epidemiologic characteristics: both are wild-zoonoses that are transmitted via a tick vector and, as such, will be driven by similar patterns of land use and land cover. This strongly supports the idea that associations with land cover pattern may offer insight into the epidemiology of a disease, at least with respect to tick-borne zoonoses. Looking at their associations with specific pattern metrics, we see that a high mean perimeter-to-area ratio is associated with higher rates of both RMSF and LYME; conversely, their associations with other measures of patch shape complexity are all negative. Since, the perimeter-to-area ratio is a function of both patch size and patch shape complexity, whereas other shape metrics are driven less by patch size, one possible explanation is that small but regularly shaped patches might favor these tick-borne infections. Our limited spatial resolution could also be factor here: given the 300m resolution of our land cover data, shape complexity of small patches would not be well captured (*e.g.* regardless of its shape, any patch of less than 90,000m^2 would be appear as a single square pixel), whereas the nuances of the shapes of larger patches would be represented more accurately. Further analyses with higher resolution land cover maps may be necessary to tease out these factors.

West Nile was positively associated with all shape metrics, except perimeter-to-area ratio. As above, this disparity may reflect the effect that patch size has on perimeter-to-area ratio or the effects of limited spatial resolution. Interestingly, West Nile was the disease most strongly associated with related circumscribing circle; and this strong positive correlation suggests that elongated, linear patches favor West Nile—this is also consistent with the observed positive association of West Nile with both the fractal and shape indices, and its negative association with perimeter-to-area ratio. Given the importance of birds as a reservoir of West Nile, and studies suggesting the role of bird migration in West Nile movement,[43-45] it is plausible that these elongated, linear patches correspond to bird migration corridors.

The large sample size and geographic extent of our data offered excellent power. Moreover, the use of validated land cover data, based on remote sensing satellites, suggests little opportunity for exposure misclassification. It is also worth noting that, while this is an ecological study, by using inherently group-level exposures and limiting our inferences to group level associations, we avoided cross-level inference, a common pitfall this design and thus eliminated the potential for ecological fallacy.

The relatively small number of diseases included in the analysis, most notably, the small number of zoonotic diseases, limits the generalizability of these conclusions. Moreover, given differing population dynamics and patterns of land use in different regions of the world, it is not clear whether or not our results will generalize outside of the contiguous United States. It is also unclear to what extent our results will generalize to other units of aggregation. Here we must consider the modifiable areal unit problem (MAUP):

associations based on spatially aggregated data can vary depending on the size and configuration of the units of aggregation. While we believe that, the scale of our aggregation units (*i.e.* counties) is appropriate to the scale of the underlying process of interest (*i.e.* interaction between humans and wild animals), we cannot know the extent to which our observed associations stem from our choice of aggregation unit, or from a zonation effect resulting from the shapes and configuration of counties. More research will be needed to address these shortcomings. Of note that the completeness of disease reporting may differ by county and, if reporting standards are systematically associated with landscape, this could introduce bias in our estimates of these associations. Similarly, reported cases include only diagnosed cases. Thus, another source of bias could exist if the percentage of cases that are diagnosed differs by county, and this difference is systematically associated with landscape. Finally, land cover is dynamic and should have changed in some areas both between the time the land cover and disease surveillance data were collected, and within each disease's five-year surveillance period. The result of such changes in land cover would almost certainly be non-differential exposure misclassification and, consequently, estimated associations being bias toward the null. Since only a very small proportion of the land area would be expected to have changed between the time of land cover and disease data collection, however, we expect this misclassification to be negligible.

This analysis suggests that associations with land cover pattern metrics may offer insight into the probability that a disease has a wild animal reservoir, and that shape metrics are likely to be the most useful. While more work is needed to refine this method—validation with a larger number of diseases and over a greater diversity of geographic regions will be

essential—the results suggest great promise for using associations with landscape to detect zoonoses and develop risk maps for emerging zoonoses.

II. HAND, FOOT AND MOUTH DISEASE AND ITS ASSOCIATIONS WITH LAND COVER TYPE AND PATTERN

Abstract

Large-scale hand, foot and mouth disease (HFMD) outbreaks have been observed in China during each of the five years since 2007, leading the Chinese Ministry of Health to include HFMD in its list of mandated notifiable diseases in 2008. While active research has improved our understanding of the disease, we know of no study to date that has investigated the landscape ecology of HFMD. This study sought to fill important gaps in our understanding of the ecology of HFMD by looking at land cover and land cover pattern, and their associations with HFMD incidence in China, between 2008 and 2011, inclusive. We developed a county-level dataset, including each county in mainland China, using HFMD surveillance data from the Chinese Center for Disease Control and Prevention, land cover data from GlobCover 2009, vegetation data from the Global Inventory Modeling and Mapping Studies, and elevation data from the Shuttle Radar Topography Mission's Digital Elevation Model. We estimated both univariate and multivariate associations between each of twelve landscape variables, plus population density, using quasi-Poisson regression models. Decreased elevation and vegetation density were significantly associated with increased rates of HFMD; and increased division, disaggregation, and diversity of land cover types were associated with increased rates of HFMD. Our results suggest connections between landscape and HFMD incidence that warrant further investigation, and support previous studies that have found local transmission to be more important than distant transmission.

Background

Hand, foot and mouth disease (HFMD) is a syndrome that occurs among a small proportion of people infected with non-polio enteroviruses. It is characterized by flu-like symptoms, rash on the hands and feet, oral lesions, loss of appetite, vomiting and diarrhea. Severe cases can produce neurological complications, including acute placid paralysis (AFP), cardiopulmonary complications and death. HFMD is most commonly observed among children, and enterovirus 71 (EV71) and coxsakievirus A16 (CA16) are the most commonly implicated pathogens. Transmission is thought to occur through fecal-oral routes and via respiratory droplets.[1]

Since the mid-1990's HFMD epidemics have become regular in the Asia-Pacific region and may be increasing in size and severity. Large-scale outbreaks of neurologically complicated HFMD have been reported in Malaysia (in 1997, 2000, 2003 and 2005), Taiwan (1998, 2000 and 2001), Australia (1999 and 2000), Singapore (2000 and 2006), Brunei (2006) and China (2007 onward).[1] Since 2007, China has experienced large-scale annual epidemics, prompting the Chinese Ministry of Health to add HFMD to its list of mandated notifiable diseases in 2008.[2,3] With the near-eradication of polio, EV71 is becoming an increasingly important cause of AFP, and some suggest that EV71 "has already made a bid to occupy the biological niche vacated by poliovirus."[46] The increasing size and frequency of HFMD outbreaks suggest that it is an important emerging public health problem.

Still, little is known about the ecology of HFMD. Analyses of surveillance data from the Chinese Center for Disease Control and Prevention (CCDC) reveal that the annual epidemics in China have moved across the country in a predictable wave moving roughly

from south to north with the end of winter, and several authors have found strong associations between weather and HFMD.[7,9,10,12-15] Yet we know of no study to date that has investigated the landscape ecology of HFMD and, specifically, associations between land cover and HFMD.

Land cover class, as determined by remote sensing satellite data, describes the dominant type of land cover (*e.g.* urban, deciduous forest, and water) as visible from above. Land cover can be described not only by class, but also by pattern, and a variety of land cover pattern metrics exist, each quantifying some aspect of land cover pattern: contagion, for example, quantifies the degree to which areas of the same land cover class are aggregated or dispersed in a landscape. Within the context of pattern metrics, a patch is a contiguous area of a single land cover type (*e.g.* a contiguous urban patch) and a landscape is the total land area included in an analysis and comprises multiple patches and classes.[47,48] Finally, land cover can be described by derived indices such as the Normalized Difference Vegetation Index (NDVI), which quantifies the density of green vegetation based on the ratio of near-infrared and visible-red intensities.[49]

This study seeks to fill important gaps in our understanding of the ecology of HFMD by looking at land cover and land cover pattern, and their associations with HFMD incidence in China. Specifically, we derive county-level estimates of elevation, land cover type, land cover pattern and vegetation density, and estimate their associations with county-level HFMD incidence for each county in mainland China, for 2008 through 2011.

Methods

Data Sources and Processing

The CCDC provided data on all recorded HFMD cases from national surveillance records, based on mandatory reporting by hospitals and clinics, for each county in China between 2008 and 2011, inclusive. Cases were diagnosed clinically or by laboratory confirmation. Reporting was via a real-time, internet-based system, with coverage exceeding 90% of county-level hospitals and 80% of clinics.[7] These data included information on case sex, age, viral type (when tested), year of HFMD onset, case severity, and case fatalities. With this, the CCDC also provided a county boundary shapefile and county-level demographic data. For this study, we restricted cases to those occurring in mainland China, and linked case data to landscape data using China's National Standard (Guo Biao (GB)) county codes. Counties, and the cases reported within them, were excluded if their GB codes could not be matched to those of known counties. Likewise, counties for which demographic data were unavailable were also excluded.

County-level land cover type and pattern variables were derived from the GlobCover 2009 land cover dataset, a global 300 meter classified land cover map with 22 land cover classes.[50] Separate county-level land cover raster files were extracted from the GlobCover dataset and analyzed to calculate 1) the proportion of each county's land area occupied by water, urban cover, agricultural cover, and snow and ice, respectively; and, 2) selected land cover pattern metrics to quantify aspects of land cover pattern including shape complexity, aggregation/disaggregation, and land cover diversity (Table 2.1). For metrics that describe pattern at the patch-level (*e.g.* shape index), we used the mean value of that metric for all

Table 2.1: Land cover classes and their assigned values under the two-class scheme

Value	GlobCover class definition [50]	Two-classes
11	Post-flooding or irrigated croplands	Developed
14	Rainfed croplands	Developed
20	Mosaic Cropland (50-70%) / Vegetation (grassland, shrubland, forest) (20-50%)	Developed
30	Mosaic Vegetation (grassland, shrubland, forest) (50-70%) / Cropland (20-50%)	Undeveloped
40	Closed to open (>15%) broadleaved evergreen and/or semi-deciduous forest (>5m)	Undeveloped
50	Closed (>40%) broadleaved deciduous forest (>5m)	Undeveloped
60	Open (15-40%) broadleaved deciduous forest (>5m)	Undeveloped
70	Closed (>40%) needleleaved evergreen forest (>5m)	Undeveloped
90	Open (15-40%) needleleaved deciduous or evergreen forest (>5m)	Undeveloped
100	Closed to open (>15%) mixed broadleaved and needleleaved forest (>5m)	Undeveloped
110	Mosaic Forest/Shrubland (50-70%) / Grassland (20-50%)	Undeveloped
120	Mosaic Grassland (50-70%) / Forest/Shrubland (20-50%)	Undeveloped
130	Closed to open (>15%) shrubland (<5m)	Undeveloped
140	Closed to open (>15%) grassland	Undeveloped
150	Sparse (>15%) vegetation (woody vegetation, shrubs, grassland)	Undeveloped
160	Closed (>40%) broadleaved forest regularly flooded - Fresh water	Undeveloped
170	Closed (>40%) broadleaved semi-deciduous and/or evergreen forest regularly flooded - Saline water	Undeveloped
180	Closed to open (>15%) vegetation (grassland, shrubland, woody vegetation) on regularly flooded or waterlogged soil - Fresh, brackish or saline water	Undeveloped
190	Artificial surfaces and associated areas (urban areas >50%)	Developed
200	Bare areas	Undeveloped
210	Water bodies	Undeveloped
220	Permanent snow and ice	Undeveloped

patches in a given county to derive the county-level value of that metric. We calculated all land cover pattern metrics twice: first, with land cover categorized in the original 22 classes; and, second, with land cover categories collapsed into only two classes, the first class corresponding to all human developed land cover types (*e.g.* urban and agricultural) and the second class corresponding to all undeveloped or natural land cover types (*e.g.* forest, grasslands, and permanent ice and snow). Under the native classification scheme (*i.e.* the 22

classes described above) results would describe associations between disease incidence and overall landscape pattern and complexity. Under a simplified scheme in which all natural classes were combined and all man-made classes were combined, results would offer insight into how the relationship between natural and developed environments influence disease risk. The land cover types under these two classification schemes are outlined in Table 2.2.

NDVI data were based on the NDVI from the Global Inventory Modeling and Mapping Studies (GIMMS), published by the Global Land Cover Facility (University of Maryland, College Park, Maryland)[51] and derived by taking a maximum value composite[49] of the six semimonthly datasets covering the summer of 2006, the most recent year included in the GIMMS data. That is, a composite dataset was created wherein each pixel's value was set equal to the maximum value of that pixel from each of the six semimonthly datasets from June through August of 2006. The maximum-value composite produces a map in which each location's peak vegetation density is used to define its NDVI, thereby eliminating the effect of seasonal and short-term fluctuations (e.g. lower vegetation density in the winter and during dry months). County-level NDVI values were then estimated separately for each county as the mean NDVI value for all non-water pixels in that county.

Elevation data were from the Shuttle Radar Topography Mission's (SRTM) Digital Elevation Model (DEM), produced by the U.S. Geological Survey (USGS).[52] The county-level elevation variable was derived by taking, for each county, the mean value of elevation from all pixels falling within that county.

Table 2.2: Summary of land cover variables, including the theoretical range of values that each metric can take, and brief descriptions.

Variable	Range	Description
Land cover type		
Normalized Difference Vegetation Index (NDVI)	0 – 1	Quantifies the density of green vegetation based on the ratio of near-infrared and visible-red intensities. Higher values indicate greater vegetation density.
Percent agriculture (%)	0 – 100	Percentage of a county covered by primarily agricultural classes (values 11, 14 & 20 in Table 2.1)
Percent ice and snow (%)	0 – 100	Percentage of a county covered by permanent ice and snow (value 220 in Table 2.1)
Percent urban (%)	0 – 100	Percentage of a county covered by primarily urban land cover (value 190 in Table 2.1)
Percent water (%)	0 – 100	Percentage of a county covered by open water (value 210 in Table 2.1)
Land cover pattern		
Contagion (CONTAG)	0 – 100	Quantifies aggregation and dispersion based on the degree to which adjacencies in a landscape are between pixels of the same or differing classes. High values indicate that pixels of the same class are clustered together; lower values indicate that pixels of the same class are dispersed throughout the landscape and interspersed with pixels of different classes.
Correlation length	≥ 0	Quantifies the connectedness, of a landscape. It is the average distance one can travel while remaining in a given patch, for all points in the landscape.
Patch density (PD)	> 0	The number of patches per hectare of landscape; quantifies landscape division.
Simpson's diversity index (SIDI)	0 - 1	"SIDI equals 1 minus the sum, across all patch types, of the proportional abundance if each patch type squared" and it "represents the probability that any 2 pixels selected at random would be different patch types." [48]
Perimeter-area ratio (PARA)	> 0	The mean perimeter-to-area ratio (m/m^2) of all patches in a landscape. Higher values may indicate either greater shape complexity or smaller patch-size.
Shape index (SHAPE)	≥ 1	A modification version of PARA that quantifies shape complexity but is unaffected by patch size. For each patch It is calculated as a quarter of the perimeter divided by the square root of the area. A value of one indicates a square patch; higher values indicate greater shape complexity. Calculated as the mean shape index for all patches in a county.

Descriptive analysis

The case population was described by determining the number and percentage of cases falling into each category of sex (male or female), age (<1, 1-1.9, 2-2.9, 3-3.9, 4-4.9, 5-9.9, or ≥10 years of age), causal virus type (CA16, EV71, or other), year of onset, and geographic region (central north, central south, central west, north east, south, south west, or west, as defined by Wang et al[7]), separately by case severity (case, severe case, or death). Incidence rates (IR) within each aforementioned category were calculated as the total number of cases reported within that category during the four-year reporting period, divided by the number in the underlying population within that category, times four-years:

$$IR_i = \frac{\sum_{y=2008}^{2011} n_{iy}}{N_i \cdot 4\,years} \tag{2.1}$$

Where n_{iy} represents the number of cases in category i and year y, and N_i represents the size of the underlying population in category i. Mortality rates (MR) were calculated similarly, but with the total number of HFMD deaths divided by person-time. Case fatality (CF) was calculated within each category, i, as the number of deaths in category i, divided by the number of cases in category i. Similarly, the mean, standard deviation, and range of each landscape variable were calculated to describe their distributions. Finally, a county-level choropleth map was created to visualize the spatial distribution of HFMD incidence rates.

Regression modeling

We fitted quasi-Poisson regression models to estimate associations between land cover variables and HFMD incidence rates at the county-level. The quasi-Poisson model was used to accommodate overdispersion in the distribution of county-level case counts. We first fit crude univariate models to separately assess the strength of the association between each land cover variable and HFMD incidence; and, then multivariate models to consider the joint associations between land cover variables and HFMD incidence. Population density was also included in our models as we felt that it could act as a confounder: population density should drive landscape (*i.e.* increasing population density would likely yield a larger proportion of urban land cover, lower NDVI, and greater disaggregation of natural land covers) and population density has been associated with HFMD incidence in previous studies.[14,53,54] All models were repeated, first using pattern metrics calculated under the two-class scheme, and, second using pattern metrics calculated under the 22-class scheme.

Normalized incidence rate ratios (NIRR) were calculated to improve the interpretability and comparability of associations with different land cover variables: for land cover variable, v, the corresponding NIRR was calculated as the coefficient of the regression model, times the standard deviation, σ, of that metric, exponentiated:

$$NIRR_v = e^{\beta_v \cdot \sigma_v} \tag{2.2}$$

Thus, each NIRR represents the ratio of the expected incidence rates in two counties having values of a given land cover variable differing by one standard deviation.

ArcGIS version 10.0 (ESRI, Redlands, CA) was used for geospatial data management tasks and mapping, with scripting in Python 2.6. Pattern metrics were calculated with FragStats 4.1 (University of Massachusetts, Amherst, MA). R version 2.12.0 was used for regression modeling, and Stata version 11.2 (StataCorp, College Station, Texas) was used for general data management and descriptive analyses.

Results

A total of 5,225,804 cases were recorded in 3,375 counties during the four-year study period. Of these, 86,362 cases (1.65%) and 26 counties (0.77%) occurred outside of mainland China and were, consequently, not included in the analysis. Records with erroneous county codes, corresponding to 438 county codes (12.98%) containing 12,120 cases (0.23%), were also excluded. Finally, demographic data were unavailable for 25 counties (0.74%) containing 17,721 cases (0.34%). The remaining 5,109,601 cases (97.78%) from 2,886 counties (85.51%) were included in the analysis. Included and excluded cases were similar in terms of age, sex, virus type, and case severity; however, the distribution of year of onset differed between, with excluded cases being more likely to have occurred in 2011. The presence of valid province codes for 97.4% of excluded cases also allowed for the comparison of spatial distribution of included and excluded cases; and excluded cases were more likely to be from south China than were included cases.

Cases were predominately males (63.0%), and 82.5% were under four-years of age. While the highest IRs and MRs were observed among children between 1 and 2 years of age, CF was highest among those under 1 year of age and declined with increasing age. Viral type

Table 2.3: Characteristics of all cases, severe cases, and fatal cases of hand, foot and mouth disease (HFMD) in mainland China, between 2008 and 2011. For each level of a characteristic, the number of cases (N), percentages of cases (among those with a known value of that characteristic)(%), and the incidence rate IR) are given. Mortality rates (MR) and case fatalities (CF) are also given, where CF is the proportion of all cases that resulted in death. IRs are given per 100,000 person-years, MRs are given per 1,000,000 person-years, and CFs are given per 100,000 cases.

	All Cases		Severe Cases		Deaths		
Characteristic	N (%)	IR	N (%)	IR	N (%)	MR	CF
Total	5,109,601 (97.8)	99.4	63,367 (1.2)	1.23	1,875 (0.04)	0.36	36.7
Age (years)							
<1.0	750,917 (14.7)	1127.0	12,933 (20.4)	19.41	549 (29.3)	8.24	73.1
1.0-1.9	1,481,047 (29.0)	2312.8	24,483 (38.6)	38.23	751 (40.1)	11.73	50.7
2.0-2.9	1,159,033 (22.7)	1844.9	13,385 (21.1)	21.31	370 (19.7)	5.89	31.9
3.0-3.9	823,380 (16.1)	1329.1	7,117 (11.2)	11.49	136 (7.25)	2.20	16.5
4.0-4.9	429,890 (8.4)	693.5	2,958 (4.7)	4.77	46 (2.45)	0.74	10.7
5.0-9.9	394,849 (7.7)	136.7	2,211 (3.5)	0.77	22 (1.17)	0.08	5.6
10+	70,471 (1.4)	1.6	278 (0.4)	0.01	1 (0.05)	0.00	1.4
Unknown	14 –	–	– –	–	– –	–	–
Sex							
Female	1,893,064 (37.0)	75.6	22,206 (35.0)	0.89	641 (34.2)	0.26	33.9
Male	3,216,535 (63.0)	121.9	41,161 (65.0)	1.56	1,234 (65.8)	0.47	38.4
Unknown	2 –	–	– –	–	– –	–	–
Virus type							
Coxsakievirus A16	43,801 (27.2)	–	1,212 (4.6)	–	18 (1.41)	–	41.1
Enterovirus 71	82,250 (51.0)	–	21,240 (81.2)	–	1,179 (92.3)	–	1433.4
Other	35,135 (21.8)	–	3,707 (14.2)	–	81 (6.34)	–	230.5
Unknown	4,948,415 –	–	– –	–	– –	–	–
Year of onset							
2008	480,861 (9.4)	37.4	2,897 (4.6)	0.23	123 (6.56)	0.10	25.6
2009	1,141,174 (22.3)	88.8	13,763 (21.7)	1.07	342 (18.2)	0.27	30.0
2010	1,926,316 (37.7)	149.8	29,330 (46.3)	2.28	936 (49.9)	0.73	48.6
2011	1,561,250 (30.6)	121.4	17,377 (27.4)	1.35	474 (25.3)	0.37	30.4
Region							
Central North	1,323,474 (25.9)	104.1	33,719 (53.2)	2.65	404 (21.5)	0.32	30.5
Central South	1,596,156 (31.2)	108.2	14,007 (22.1)	0.95	555 (29.6)	0.38	34.8
Central West	279,593 (5.5)	74.4	1,917 (3.0)	0.51	115 (6.13)	0.31	41.1
North East	280,697 (5.5)	64.7	1,863 (2.9)	0.43	90 (4.80)	0.21	32.1
South	1,224,281 (24.0)	171.9	7,258 (11.5)	1.02	435 (23.2)	0.61	35.5
South West	373,467 (7.3)	48.6	4,532 (7.2)	0.59	270 (14.4)	0.35	72.3
West	31,933 (0.6)	30.5	71 (0.1)	0.07	6 (0.32)	0.06	18.8

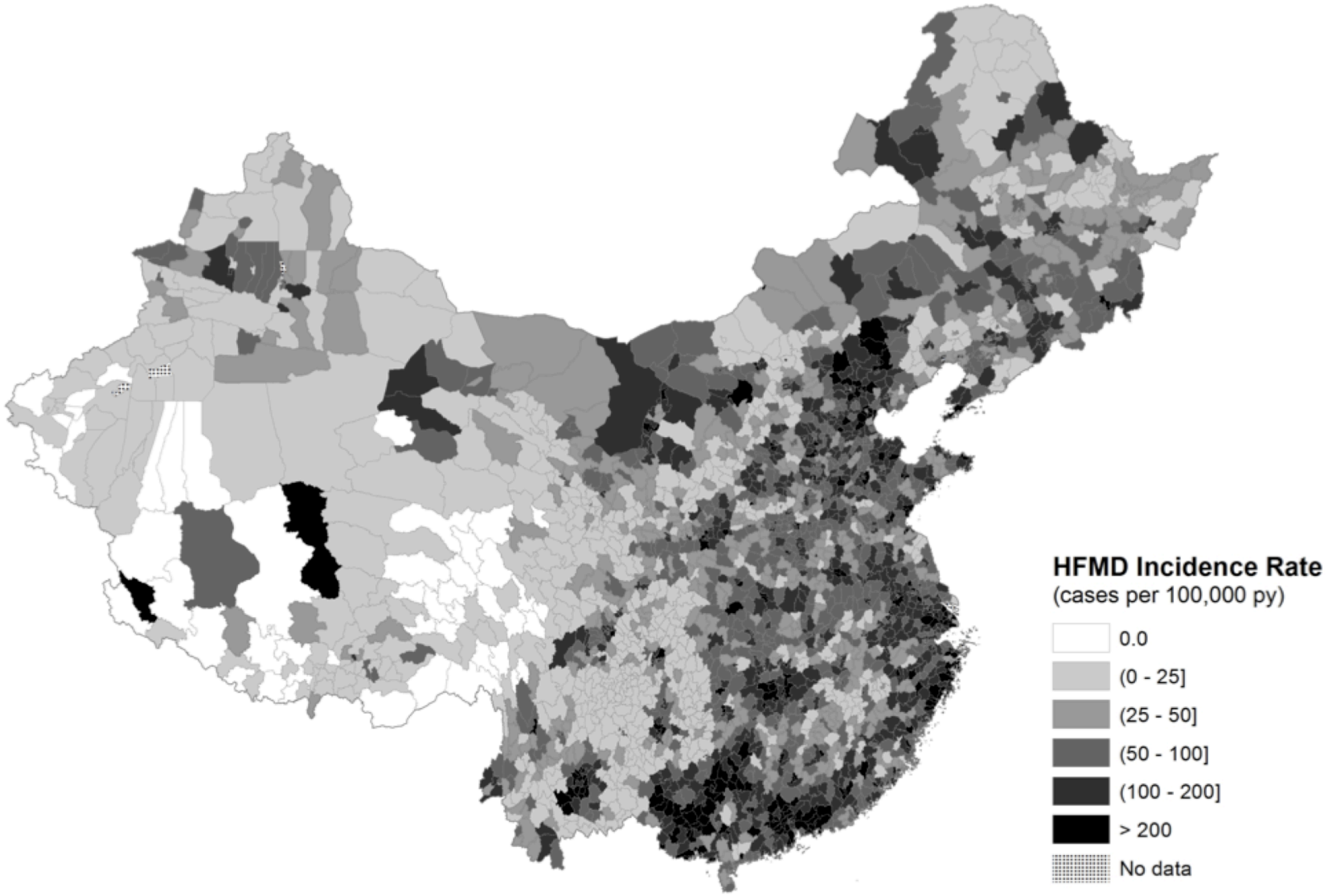

Figure 2.1: County-level incidence rates of hand, foot and mouth disease (HFMD) for all counties in mainland China, between 2008 and 2011, inclusive.

was determined for 3.2% of cases (*n*=161,186), of whom the majority had EV71 (51.0%); moreover, EV71 was responsible for a disproportionate fraction of severe cases (81.2%) and deaths (92.3%), suggesting it to be the more virulent causal agent. Both incidence and CF appear to have peaked in 2010. IRs were highest in the central south region, while CF was highest in the south west region (Table 2.3 and Figure 2.1).

The mean percentage of each county's land covered by agricultural, ice and snow, urban, and water were 46.0%, 0.3%, 6.6%, and 1.9%, respectively. Counties ranged from very sparsely, to very densely vegetated (NDVI range: 0.04, 0.92), but had moderate vegetation density, on average (mean NDVI: 0.58). The mean elevation was 757 meters (SD = 1,041), with counties ranging from less than one-meter to 5,152 meters above sea level. Population density ranged from less than one person per square-kilometer (km^2) to 53,136 people per km^2 (Table 2.4).

Elevation was negatively associated with HFMD incidence, with a 1σ increase in elevation being associated with a 41% to 43% lower HFMD incidence, depending on the model. Vegetation density was also negatively associated with HFMD, with each 1σ increase in NDVI being associated with a roughly 20% lower IR. Each 1σ increase in the proportion of urban land cover was associated with a 12% to 19% greater incidence of HFMD. The percentage of a county's surface covered by open water was significantly positively associated with HFMD in the univariate models; however these associations were attenuated and no longer significant in the multivariate models. In the univariate analysis, higher population densities were associated with higher rates of HFMD; however the direction of this association reversed when other variables were included in the model.

Table 2.4: Summary of the distributions of landscape exposure variables including, for each variable, the mean, standard deviation (SD), and minimum and maximum values (Range). The top portion of the table includes non-pattern metrics; the center and bottom portions include pattern metrics calculated under the two-class and 22-class schemes, respectively.

Variable	Mean	(SD)	Range
Elevation (meters)	757	(1,040)	0.00 – 5,152
Normalized Difference Vegetation Index	0.58	(0.15)	0.04 – 0.92
Percent agriculture (%)	46.0	(30.5)	0.00 – 99.0
Percent ice & snow (%)	0.26	(1.92)	0.00 – 52.4
Percent urban (%)	6.62	(16.3)	0.00 – 100
Percent water (%)	1.94	(4.75)	0.00 – 70.0
Population density (population per km^2)	1176	(3,620)	0.09 – 53,136
Pattern metrics, two-classes			
Contagion (%)	44.6	(26.2)	0.00 – 99.9
Correlation length (meters)	14,170	(13,830)	575 – 175,200
Perimeter-to-area ratio (m/m^2)	100	(15.2)	0.45 – 123
Patch density (patches per hectare)	0.17	(0.10)	0.00 – 1.19
Shape index	1.32	(0.11)	1.03 – 2.52
Simpson's diversity index	0.29	(0.16)	0.00 – 0.50
Pattern metrics, twenty-two-classes			
Contagion (%)	51.7	(14.9)	0.00 – 99.1
Correlation length (meters)	7,312	(9,868)	409 – 139,900
Perimeter-to-area ratio (m/m^2)	106	(7.85)	10.9 – 118
Patch density (patches per hectare)	0.90	(0.46)	0.01 – 3.01
Shape index	1.23	(0.05)	1.05 – 1.60
Simpson's diversity index	0.63	(0.20)	0.00 – 0.90

Under both classification schemes, increased land cover diversity, as measured by Simpson's diversity index, was associated with increased HFMD incidence in the univariate analyses; none of these associations maintained significance in the multivariate models. Similarly, increased disaggregation and dispersion of land cover, as measured by contagion, were associated with increased HFMD in the univariate models; but these associations failed to maintain significance in the multivariate models. Increasing landscape division, as

measured by patch density, was associated with increased HFMD incidence in all but the multivariate 22-class model; these associations were most pronounced under the two-class scheme, with NIRRs ranging from 1.07 to 1.20. Increased connectedness, as measured by correlation length, was consistently and significantly associated with lower rates of HFMD. Associations between patch shape metrics (*i.e.* the shape index, and the perimeter-to-area ratio) and HFMD were less consistent. Under the two-class scheme, increased patch shape complexity (shape index) was associated with increased HFMD, but only in the univariate analyses; no clear picture of associations with perimeter-to-area ratio was present. Under the 22-class scheme, shape index, but not perimeter-to-area ratio was slightly negatively associated with HFMD incidence (Table 2.5).

Discussion

Our results suggest that associations exist between several elements of landscape and HFMD incidence. Our finding that higher elevation is associated with less HFMD is unsurprising: several studies have noted increased HFMD occurring with warmer temperatures[7,9,10,12-15] and, given the strong effect of elevation on temperature, it is likely that elevation is acting as a crude proxy for temperature. The effect of elevation may also be mediated by other components of weather, including humidity, which is both positively associated with HFMD and negatively associated with elevation. Conversely, in the association observed between vegetation density (*i.e.* NDVI) and HFMD, it seems very unlikely that NDVI is acting as a proxy for weather: warm, humid weather favors both high vegetation density and high HFMD and, if NDVI were simply acting as a proxy for weather, we would expect a positive

Table 2.5: Normalized incidence rate ratios (NIRR) and 95% confidence intervals for associations between landscape variables and hand, foot and mouth disease (HFMD) incidence in mainland China, between 2008 and 2011. For each variable, NIRRs are given under univariate and multivariate parameterization of quasi-Poisson models. Univariate NIRRs for non-pattern variables (*i.e.* elevation, Normalized Difference Vegetation Index (NDVI), population density, and percent agriculture, ice and snow, urban and water) do not vary between the two-class and 22-class schemes, and are repeated for convenience.

Metric	Univariate	Multivariate
Two-Classes		
Elevation	0.57 (0.53, 0.62) *	0.59 (0.54, 0.64) *
NDVI	0.81 (0.78, 0.83) *	0.81 (0.78, 0.84) *
Percent agriculture	0.94 (0.91, 0.98) *	0.86 (0.82, 0.89) *
Percent ice & snow	0.50 (0.25, 1.01)	0.93 (0.68, 1.26)
Percent urban	1.15 (1.13, 1.18) *	1.12 (1.06, 1.19) *
Percent water	1.13 (1.10, 1.15) *	1.02 (0.99, 1.05)
Population density	1.07 (1.04, 1.09) *	0.78 (0.74, 0.83) *
Contagion	0.95 (0.92, 0.98) *	1.02 (0.89, 1.17)
Correlation length	0.48 (0.43, 0.52) *	0.60 (0.54, 0.67) *
Perimeter-to-area ratio	1.02 (0.98, 1.06)	1.07 (1.01, 1.13) *
Patch density	1.20 (1.16, 1.24) *	1.07 (1.02, 1.12) *
Shape index	1.05 (1.02, 1.09) *	0.96 (0.92, 1.01)
Simpson's diversity index	1.05 (1.01, 1.08) *	0.98 (0.87, 1.11)
Twenty-two-Classes		
Elevation	0.57 (0.53, 0.62) *	0.59 (0.55, 0.64) *
NDVI	0.81 (0.78, 0.83) *	0.80 (0.77, 0.83) *
Percent agriculture	0.94 (0.91, 0.98) *	1.02 (0.97, 1.06)
Percent ice & snow	0.50 (0.25, 1.01)	0.87 (0.64, 1.18)
Percent urban	1.15 (1.13, 1.18) *	1.19 (1.13, 1.27) *
Percent water	1.13 (1.10, 1.15) *	1.01 (0.98, 1.03)
Population density	1.07 (1.04, 1.09) *	0.81 (0.76, 0.86) *
Contagion	0.89 (0.86, 0.92) *	0.94 (0.84, 1.04)
Correlation length	0.54 (0.50, 0.60) *	0.70 (0.62, 0.78) *
Perimeter-to-area ratio	1.01 (0.98, 1.04)	0.98 (0.93, 1.03)
Patch density	1.16 (1.12, 1.20) *	0.98 (0.92, 1.06)
Shape index	0.91 (0.88, 0.94) *	0.86 (0.82, 0.90) *
Simpson's diversity index	1.12 (1.08, 1.16) *	1.09 (0.99, 1.20)

* Indicates that the 95% confidence interval excludes 1.0

association between NDVI and HFMD. Instead, we observed a consistent and reasonably strong negative association between the two. It is also unlikely that this association results from confounding by population density, since the association remained nearly unchanged in the multivariate models that included population density and percent urban land cover.

Our findings with regard to population density are interesting. Previous studies have differed with regard to the association between population density and HFMD: Hu et al[14] and Zhu et al[54] found positive associations between population density and HFMD incidence; while Qiaoyun et al[53] reported the greatest risk of HFMD outside of densely populated city centers. As with Hu and Zhu, our univariate models found a positive association between population density and HFMD. However, once adjusted for landscape variables in our multivariate analyses, we observed a significant negative association between population density and HFMD. This suggests that the positive univariate association may actually be driven by elements of landscape, most notably, elevation (higher population densities and higher rates of HFMD both occur at lower elevations) and percent urban land cover (higher population densities and higher rates of HFMD both occur in counties with more urban land cover). As with Qiaoyun, therefore, our findings suggest that sprawl and suburban landscapes—those with a great deal of developed and artificial land surface but lower population densities—may be those most favorable to HFMD. HFMD risk may be lower in the most densely populated city centers, since improved public health infrastructures, and water supply and sanitation systems, for example, may be effectively reducing risk in these places.

Positive associations with patch density and Simpson's diversity index suggest that diverse and divided land cover favors higher HFMD rates. That the effect of patch density was more pronounced under the two-class scheme suggests that anthropogenic land cover division is more relevant than natural division. Similarly, the consistent and statistically significant associations seen between correlation length and HFMD suggest that HFMD occurs at greater rates in counties with disconnected patches. The underlying mechanism for these associations is not entirely clear. That said, if we believe that the observed large-scale patterns of spread (movement from south to north in the spring, for example) result from general diffusion of infections throughout populations, then these results do seem surprising. Research in landscape ecology suggests that heterogeneous and fragmented landscapes should reduce the spread of disturbances (including "species-specific parasites") that are restricted to a single habitat type, in this case developed or urban environments.[55,56] Our results, then, suggest that HFMD occurs most often in those counties with landscapes that present the greatest resistance to spread via broad population diffusion or percolation. This, in turn, suggests that the movement of HFMD epidemics may occur through introduction events in which an index cases initiates a seasonal outbreak in a community; or, alternatively, that the causal pathogens may maintain a broad distribution in environmental reservoirs but only produce outbreaks in those times and places when the weather is most favorable. Wang et al[7] attempted to parse transmission into "human-to-human within prefecture", "human-to-human between prefecture" and "reservoir-to-human", and found transmission within prefectures to dominate the later two types. This is consistent with the

observed associations with landscape, which also suggest that HFMD transmission is primarily local.

It is possible that the associations observed in our county-level analysis would differ at other levels of spatial aggregation and we, therefore, cannot draw any conclusions about the influence of local landscape on an individual's HFMD risk. It is also possible that ecology would be better described for each causal agent separately (*i.e.* EV71 and CA16) than for HFMD as a whole; however, we felt that the small number of cases with a known viral type, and the resulting unstable virus-specific county-level estimates, would not have supported a meaningful and trustworthy analysis. Moreover, as our data were limited to reported cases of HFMD, we have no information on cases that didn't result in medical care and consequent capture by surveillance. If spatial patterns of health care access or utilization are somehow associated with our landscape, then our estimates of the associations between landscape and HFMD may be biased. Without information on county-level access and utilization, however, the presence and exact nature of such a bias would be difficult to predict. Conversely, given our use of established and validated sources of land cover data, we see little opportunity for exposure misclassification. Moreover, by analyzing data across all of mainland China, our results should be more broadly generalizable than if we had restricted our analysis to a smaller or more ecologically homogeneous area. Finally, the large number of cases in our dataset offered excellent power and allowed us to include all of our predictor variables in a single multivariate model with very little chance of over-fitting.

Our results suggest connections between landscape and HFMD incidence. We found that HFMD rates are highest in counties at lower elevations, with less dense vegetation, a

greater proportion of urban landscape, lower population densities, and more diverse and fragmented land cover. The mechanisms underlying some of these associations are unclear and raise questions that may be useful in informing future research.

III. The role of climate change in the growth of hand, foot and mouth disease in China

Abstract

Outbreaks of hand, foot and mouth disease (HFMD) have become increasingly regular in the Asia-Pacific region, and China has experienced annual epidemics each year since 2007. Several studies have found associations between weather and HFMD, suggesting that climate change could have a role in the recent growth of HFMD in China. We sought to determine if climate change could underlie the recent emergence and growth of HFMD in China by developing a weather-based predictive model of HFMD and applying that model to historical climate data. When monthly climate-based HFMD predictions were regressed against calendar time, we found evidence of a significant increasing secular trend, with predicted rates for 2011 being 94% higher than those for 1982 (Incidence rate ratio (IRR): 1.937; 95% confidence interval (CI): 1.933, 1.940). Most of the increase in the predicted HFMD incidence occurred between 2002 and 2011, with predicted rates for 2011 being 49% higher than those for 2001 (IRR = 1.490; 95% CI: 1.488,1.493). Our climate-based retrospective predictions suggest that changing climate should have made weather increasingly favorable to HFMD during our thirty-year study period and we find that the data are compatible with climate change playing a role in the recent growth of HFMD in China.

Background

Hand, foot and mouth disease (HFMD) occurs in some cases of infection with non-polio enteroviruses, most notably infections with enterovirus 71 (EV71) and coxsakievirus 16

(CA16) among children under 5 years of age.[1] Though historically responsible for only modest and sporadic outbreaks, since the mid-1990's epidemics have become regular in the Asia-Pacific region and appear to be increasing in size and severity. In that time, large-scale outbreaks of neurologically complicated HFMD have occurred in Malaysia (in 1997, 2000, 2003 and 2005), Taiwan (1998, 2000 and 2001), Australia (1999 and 2000), Singapore (2000 and 2006), Brunei (2006) and China (2007 through 2010).[1] China has experienced annual epidemics each year since 2007, prompting the Chinese Ministry of Health to add HFMD to its list of mandated notifiable diseases in 2008.[3]

In temperate regions HFMD exhibits pronounced seasonality, with the most notable peaks in incidence occurring in summer.[5-7,16-18] Seasonal variability is most dramatic at higher latitudes and, while the overall pattern of increased incidence in the summer remains, the degree of seasonal variability is less dramatic nearer the equator.[5,11,16,19] With that, a number of studies have looked at associations between meteorological variables and HFMD incidence.[5-15] Studies generally agree that HFMD incidence is positively associated with higher temperatures, though there may be a threshold temperature above which transmission begins to decline. Less dramatically, there appears to be a modest, positive association between relative humidity and HFMD incidence. Most of the studies reviewed have found some association between precipitation and HFMD incidence, but they have differed with regard to the exact nature of that association: Wang et al[7] found the greatest risk with medium to high precipitation, Ma et al[6] found a positive association between precipitation and HFMD, and Hii et al[5] found an inverted-U Association with incidence peaking after periods of moderate rain and lowest after periods of high rainfall.

These observed associations between weather and HFMD suggest that climate change could potentially explain the recently established seasonal pattern of HFMD epidemics in China. While Ma et al[6] and Urashima et al[11] both considered the potential for a connection between climate change and HFMD, we know of no study that has directly investigated what role climate change may have played, if any, in the recent growth in HFMD observed throughout the Asia-Pacific region.

With that, we sought to determine if climate change could underlie the recent emergence and growth of HFMD in China. We modeled the relationship between meteorological variables and HFMD incidence, aggregated by county and month, for years with surveillance data. We then applied this model to historical climate data, to retrospectively predict the expected national HFMD incidence based on weather in each county and during each month from 1982 through 2011. Finally, we tested for the presence of a secular temporal trend in these predictions to determine if changes in climate could underlie observed increases in HFMD in China.

Methods

Data Sources and Processing

The Chinese Center for Disease Control and Prevention (CCDC) provided data on all recorded HFMD cases from national surveillance records, based on mandatory reporting by hospitals and clinics, for each county in China during each month between 2008 and 2011, inclusive. Cases were diagnosed clinically or by laboratory confirmation. Reporting was via a real-time, internet-based system, with coverage exceeding 90% of county-level hospitals and

80% of clinics.[7] With this, the CCDC also provided a county boundary shapefile and county-level demographic data. For this study, we restricted cases to those occurring in mainland China, and linked case data to climate data on Guo Biao (GB) county codes. Counties, and the cases reported within them, were excluded if their GB codes could not be matched to those of known counties. Likewise, counties for which demographic data were unavailable were also excluded.

Climate data were derived from the National Oceanic and Atmospheric Administration's (NOAA) "Global summary of the day" (GSOD) data,[57] recorded from a global network of ground-based weather stations. We acquired data from all weather stations located in either mainland China, or within a 500 kilometer buffer surrounding mainland China. This buffer was included to ensure adequate data for interpolation near borders. Data availability was very limited prior to 1973; and, between 1973 and 1981, complete data were available from less than 8% of all stations in our study area. To ensure sufficient data for robust interpolation, we, therefore, restricted our historical climate dataset to the period from 1 January 1982 through 31 December 2011. For each weather station and each month, data were collapsed from their native daily frequency, to monthly frequency. Where a variable recorded a daily maximum (*e.g.* maximum temperature, maximum wind gust) or minimum (*e.g.* minimum temperature) we took the corresponding monthly maximum or minimum to derive the monthly value. For all other variables, we took the monthly mean. Raster surfaces for each weather variable were generated for each month using cokriging with elevation as a covariate. Elevation data were from the Shuttle Radar Topography Mission's (SRTM) Digital Elevation Model (DEM), produced by the U.S.

Geological Survey (USGS).[52] Finally, for each meteorological variable, and each month, we extracted a corresponding county-level variable by taking, for each county, the mean value of that meteorological variable from all pixels falling within that county.

Model building

We modeled the relationship between meteorological variables and HFMD incidence between 2008 and 2011, the years during which the CCDC conducted nationwide HFMD surveillance, to develop a weather-based predictive model for HFMD. We trained the model using data from 2008 through 2010, and withheld the data from 2011 as our validation set. We chose the elements of weather to include in our model based on results of earlier studies that looked at associations between weather and HFMD. These elements included temperature, humidity, wind speed, and precipitation. Multiple possible variables were available to represent temperature (mean, minimum, and maximum temperature), humidity (dew point and relative humidity) and wind speed (mean wind speed, maximum sustained wind speed and maximum wind gust). Where alternate variables were strongly correlated only one variable was included in the model to avoid redundancy and problems with collinearity, and variable selection was based on Akaike information criterion (AIC); when they were not strongly correlated, multiple alternate variables were allowed in the model. The final set of core variables comprised five: mean temperature, mean daily precipitation, mean relative humidity, maximum wind gust, and mean wind speed. Previous authors have found non-linear associations between weather parameters and HFMD[7] and we, consequently, included quadratic terms for each meteorological variable in our model. For each meteorological

variable, we then tested the effect of including its value from the prior month (*i.e.* one-month lagged variable). Lagged variables were kept in the model if their inclusion resulted in a decrease in AIC (AIC-Δ) of at least 4.0. This cut point was selected based on literature suggesting that, when comparing two nested models, an AIC-Δ of at least 4.0 indicates that evidence for the more parsimonious model is "considerably less" (AIC-Δ of 4.0-7.0) or "essentially none" (AIC-Δ of >10), relative to the larger model.[58] Similarly, we tested for the effect of including interaction terms between meteorological variables, and retained those whose inclusion resulted in an AIC-Δ of at least 4.0. Each predictor variable was centered about its mean and scaled to a neat value near its standard deviation (*e.g.* the standard deviation of temperature was 19.1°F and temperature was scaled so that its coefficient corresponded to the effect of a 20°F increase in temperature).

Model validation

We trained our model using data from 2008 through 2010, and tested its performance against data from 2011. Predictive performance was quantified by calculating the coefficient of the correlation between the number of observed and predicted cases (r) and by the mean absolute error (MAE):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|P_i - O_i| \tag{3.1}$$

Where n is the number of observations in the sample, O_i is the observed number of cases in observation i, and P_i is the model's prediction of the number of cases in observation i.

Correlation coefficients and MAEs were calculated separately for the training and validation samples, and these calculations were repeated at two levels of aggregation: first, at the county-level, which was the level of aggregation of the data used to build the model, and in which each observation contains the number of cases in a given county and month; and, second, at the national-level, in which cases in all counties have been totaled by month. Finally, we plotted the time-series of the number of cases observed nationally, by month, against the model-based predictions to visually assess model fit and predictive accuracy.

Modeling predicted change in HFMD over time

We applied the predictive model to historical climate data and estimated the expected incidence of HFMD, based on weather, during each month, and in each county, from 1 January 1982 through 31 December 2011. These county-level predictions were totaled separately for each month to create a predicted monthly, national HFMD time-series. We then fit a generalized estimating equations (GEE) Poisson regression model, with autoregressive (AR 1) correlation structure, calendar year, year-squared and indicator variables for calendar month as the independent variables, and predicted number of HFMD cases as the dependent variable. Year was centered about its mean value (1996.5) and scaled so that the resulting coefficients corresponded to estimated change in HFMD with each ten-year change in calendar date.

The resulting function defined by the coefficients for calendar year and year-squared, therefore, represents the predicted secular trend in HFMD that we would expect to have occurred based on changes in climate. In turn, a positive secular trend in these weather-based

predictions would suggest that climate has changed in a manner that is consistent with increased HFMD incidence.

Results

A total of 5,225,804 cases were recorded in 3,375 counties during the four-year surveillance period. Of these, 86,362 cases (1.65%) and 26 counties (0.77%) occurred outside of mainland China and were, consequently, not included in the analysis. Records with erroneous county codes, corresponding to 438 county codes (12.98%) containing 12,120 cases (0.23%), were also excluded. Finally, demographic data were unavailable for 25 counties (0.74%) containing 17,721 cases (0.34%). The remaining 5,109,601 cases (97.78%) from 2,886 counties (85.51%) were included in the analysis. Of these 9.4% of cases (*n*=480,861) occurred in 2008, 22.3% (*n*=1,141,174) occurred in 2009, 37.7% (*n*=1,926,316) occurred in 2010, and 30.6% (*n*=1,561,250) occurred in 2011. Monthly county-level incidence rates ranged from zero to 13,314 per 100,000 person-years (mean: 93.8; standard deviation (SD): 218.6), and monthly county-level case counts ranged from zero to 4,668 cases (mean: 36.9; SD: 101.6). Monthly national incidence rates ranged from 0.03 to 30.0 per 100,000 person-years (mean: 8.3; SD: 7.5) and monthly national case counts ranged from 369 to 386,561 (mean: 106,450; SD: 97,159). Most cases were males (63.0%) and under 5 years of age (90.9%).

Weather varied widely throughout mainland China during the study period. Mean monthly temperatures varied from -28.9° to 87.0°F, mean precipitation ranged from zero to 0.87 inches, relative humidity ranged from 7.3% to 100%, mean wind speeds ranged from 0.9 knots to 12.0 knots, and maximum monthly gust speeds ranged from 16.9 to 58.8 knots

Table 3.1: Summary of monthly county-level weather variables for mainland China from 1982 through 2011

Variable	Mean (SD)	Range
Mean temperature (°F)	55.9 (19.1)	–28.9, 87.9
Mean precipitation (inches)	0.10 (0.09)	0.00, 0.87
Mean relative humidity (%)	64.5 (14.2)	7.27, 100
Mean wind speed (knots)	4.57 (1.20)	0.90, 12.0
Maximum wind gust (knots)	37.3 (4.11)	16.9, 58.8

(Table 3.1). The period during which the CCDC conducted nationwide HFMD surveillance, 2008 through 2011, was slightly warmer, less humid, and less windy than was the preceding 26-year period.

Our final model of associations between meteorological variables and HFMD incidence included mean temperature, mean daily precipitation, mean relative humidity, maximum wind gust, mean wind speed, one-month lagged temperature, and the square of each variable. Temperature was the only lagged variable that met our criterion for inclusion in the final model (*i.e.* AIC-$\Delta \geq 4.0$); no interaction terms met our inclusion criterion. We found that HFMD risk was greatest at higher temperatures, and with lagged temperatures in the range of 50° to 55°F. Within the range of their most common values, HFMD risk was greater with higher precipitation, higher humidity, higher mean wind speeds, and lower maximum gust speeds; though the directions of these associations reversed at extremely low or high values of these predictors (Figure 3.1). All variables in the final model were significantly associated with HFMD incidence at $p<0.05$ and most were significant at $p<0.001$ (Table 3.2).

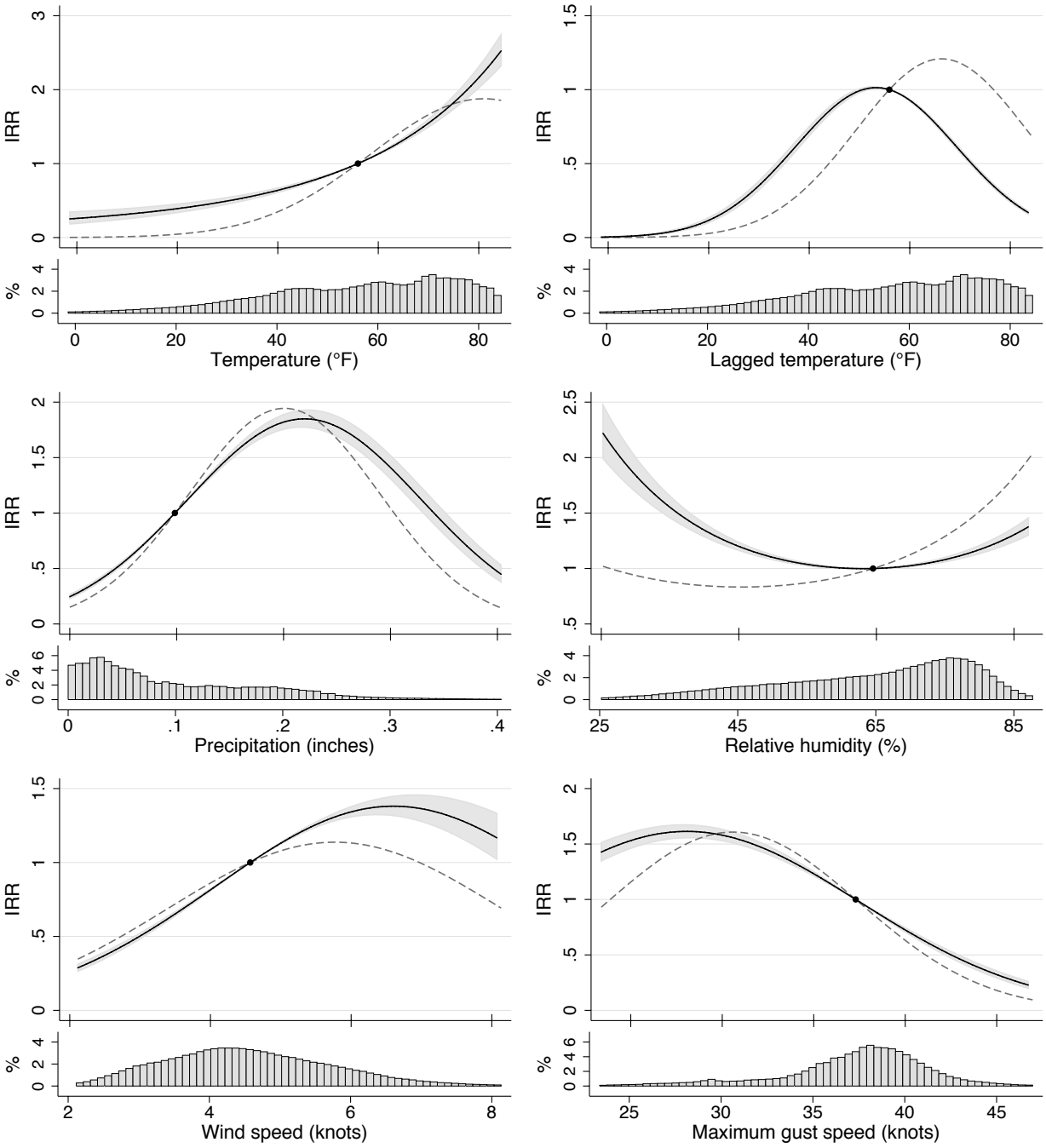

Figure 3.1: Associations between weather variables and hand, foot and mouth disease (HFMD) incidence. Lines and shaded areas show the incidence rate ratios (IRR) and their 95% confidence intervals for the association between each weather parameter and HFMD incidence, relative to that variable's mean (indicated by the small circle): dashed lines show crude, univariate associations; solid lines show associations from the final multivariate model. The lower panel of each graph shows the histogram of the distribution of the weather variable, truncated at the 0.5th and 99.5th percentiles (*i.e.* the central 99% of the distribution).

Table 3.2: Coefficients and 95% confidence intervals of the predictive model of hand, foot and mouth disease. Each variable was centered on its mean and scaled to a neat value near its standard deviation.

Predictor	Coefficient	(95% CI)
Temperature (20°F)		
Linear	0.37**	(0.18, 0.56)
Squared	0.04*	(0.01, 0.07)
Lagged, linear	4.16**	(3.97, 4.35)
Lagged, squared	-0.78**	(–0.81, –0.75)
Precipitation (0.1 inches)		
Linear	1.02**	(0.96, 1.07)
Squared	-0.42**	(–0.45, –0.39)
Relative humidity (15%)		
Linear	-1.05**	(–1.19, –0.91)
Squared	0.13**	(0.11, 0.14)
Wind speed (1 knot)		
Linear	0.32**	(0.30, 0.33)
Squared	-0.08**	(–0.09, –0.07)
Maximum gust (5 knots)		
Linear	1.57**	(1.36, 1.79)
Squared	-0.14**	(–0.16, –0.12)
Intercept	-18.23**	(–19.05, –17.42)

* $p < 0.05$
** $p < 0.001$

We found that our model was able to capture the general timing and intensity of HFMD outbreaks in both the training and validation samples (Figure 3.2). At the county-level, the MAE was 34.0 cases in the training sample, and 39.0 cases in the validation sample, with correlation coefficients of 0.54 and 0.53, respectively. At the national-level, the MAE was 45,257 cases in the training sample and 33,747 cases in the validation sample, with

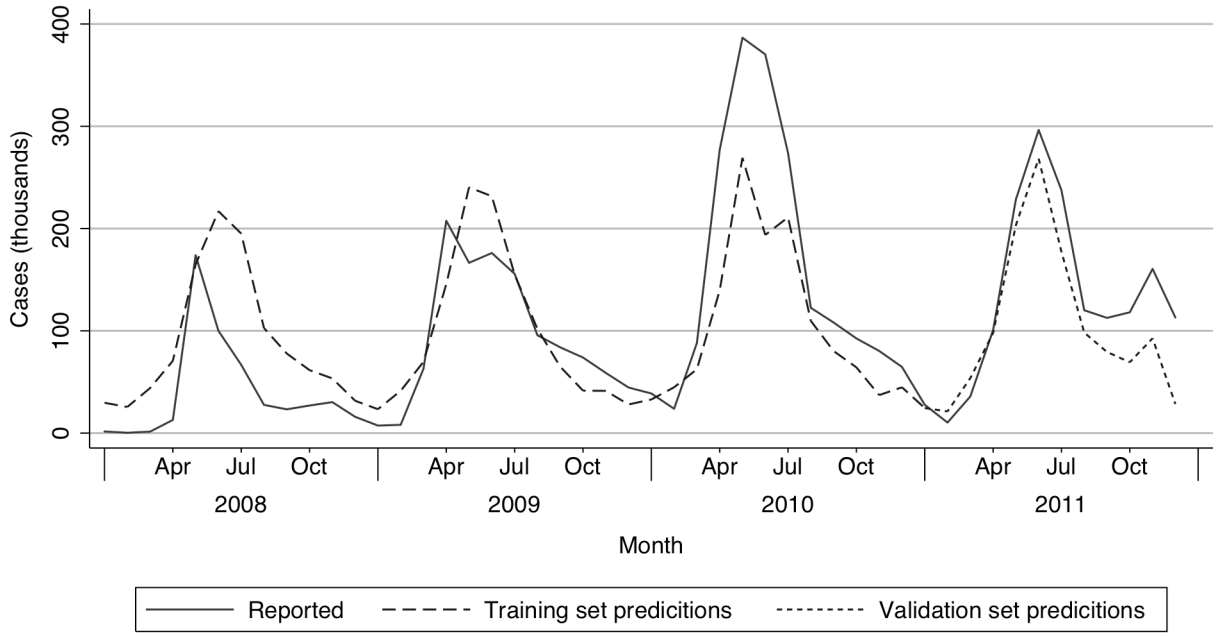

Figure 3.2: Time series of the number of cases of hand, foot and mouth disease reported each month in mainland China, between 2008 and 2011 (solid line), and the number of cases predicted each month by the weather-based model (dashed line). Predictions for 2008 through 2010 are in the same sample used to train the regression model (long dashes); those for 2011 are in the validation sample (short dashes).

correlation coefficients of 0.78 and 0.93, respectively. The similar correlation coefficients and MAEs in the training and validation sets at the county-level, and superior predictive performance in the validation set at the national-level suggests little, if any, over-fitting of our model. When applied to historical climate data, our model predicted the largest outbreak in 2010, with the next three largest in 2011, 2009 and 2008, respectively; and these predictions correspond to largest reported outbreaks based on surveillance. Our model also predicted the summertime peaks that have been widely observed with HFMD (Figure 3.3).

While the predicted incidence fluctuates from year to year, a general upward trend is apparent, especially from 2006 onward (Figure 3.3). When monthly climate-based HFMD

predictions were regressed against calendar time, we found significant positive associations with year and year-squared, relative to the middle of the predicted time-series (year 1996.5), indicating an increasing secular trend (Table 3.3), with predicted rates for 2011 being 94% higher than those for 1982 (Incidence rate ratio (IRR): 1.937; 95% confidence interval (CI): 1.933, 1.940). Predicted HFMD incidence remains low and relatively stable during the first two-decades of predictions; most of the predicted increase in HFMD incidence is seen in last ten-years, between 2002 and 2011, with predicted rates for 2011 being 49% higher than those for 2001 (IRR = 1.490; 95% CI: 1.488,1.493)(Figure 3.4).

Discussion

We found strong associations between weather and HFMD incidence in mainland China between 2008 and 2011, and developed a predictive model that successfully captured the

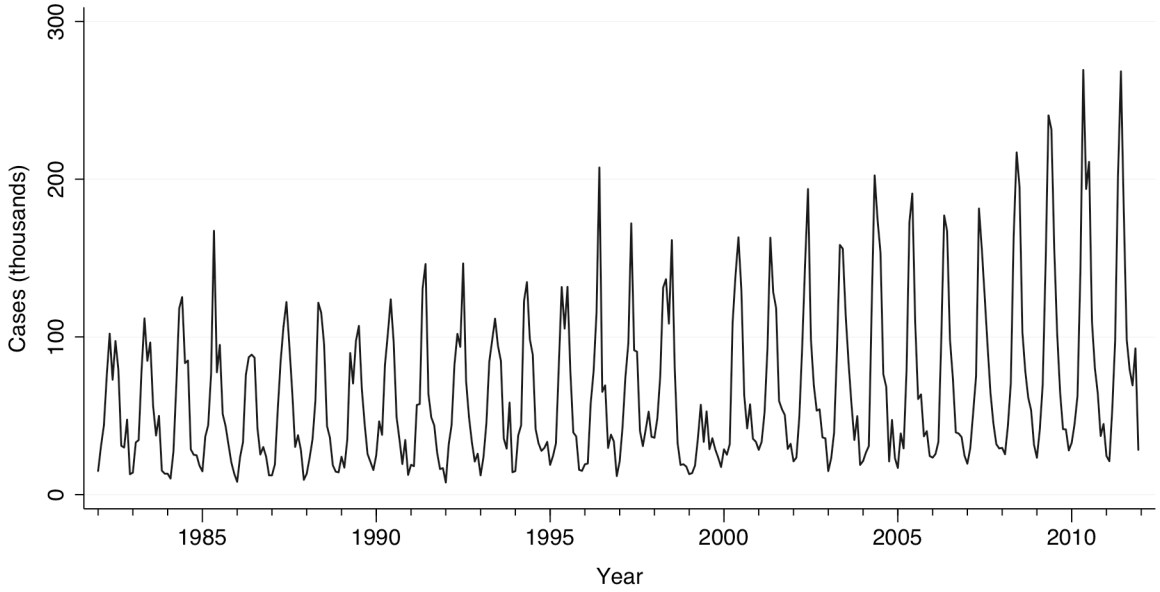

Figure 3.3: Time series of the predicted number of hand, foot and mouth disease (HFMD) cases in mainland China, during each month from January 1982 through December 2011, based on associations with weather.

Table 3.3: Incidence rate ratios (IRR) and 95% confidence intervals describing the association between calendar time and predicted hand, foot and mouth disease (HFMD) incidence. Year is centered on its mean value and scaled, so that IRRs represent the relative change in incidence associated with each ten-year change from 1996.5.

	IRR	(95% CI)
Year (per 10 years)	1.256	(1.255, 1.257)
Year (per 10 years), squared	1.114	(1.113, 1.115)
Month		
January	0.136	(0.136, 0.137)
February	0.200	(0.200, 0.201)
March	0.313	(0.312, 0.313)
April	0.625	(0.624, 0.625)
May	1.0	(Reference)
June	0.968	(0.967, 0.969)
July	0.801	(0.800, 0.802)
August	0.484	(0.484, 0.485)
September	0.324	(0.323, 0.324)
October	0.262	(0.262, 0.263)
November	0.241	(0.240, 0.241)
December	0.155	(0.155, 0.156)

general timing and intensity of HFMD outbreaks, both in the subset of data used for training and in the subset set aside for validation. Our model suggests that HFMD is favored by warmer temperatures, following moderately cool months, combined with moderately high precipitation, very high or very low relative humidity, and higher mean wind speeds with lower maximum gust speeds. These associations are consistent with the results of most prior studies of associations between weather and HFMD.

Our climate-based retrospective predictions suggest that changing climate should have produced weather increasingly favorable to HFMD during our thirty-year study period. While we cannot definitively establish climate change as the primary or partial cause of the observed increases in HFMD, our results suggest that the data are compatible with climate

change playing a role in these observed increases. Of note, our model predicted an average of 717,295 cases per year between 1982 and 2007, the period before large-scale annual outbreaks are known to have occurred. While it is likely that some HFMD occurred in China during that period, it is almost certain that our climate-based predictions severely overestimate the number of cases prior to 2007. If climate change has been the primary driver of increased HFMD in China, then this suggests that our model failed to capture some aspect of the connection between climate and HFMD: whether that is some aspect of weather not included

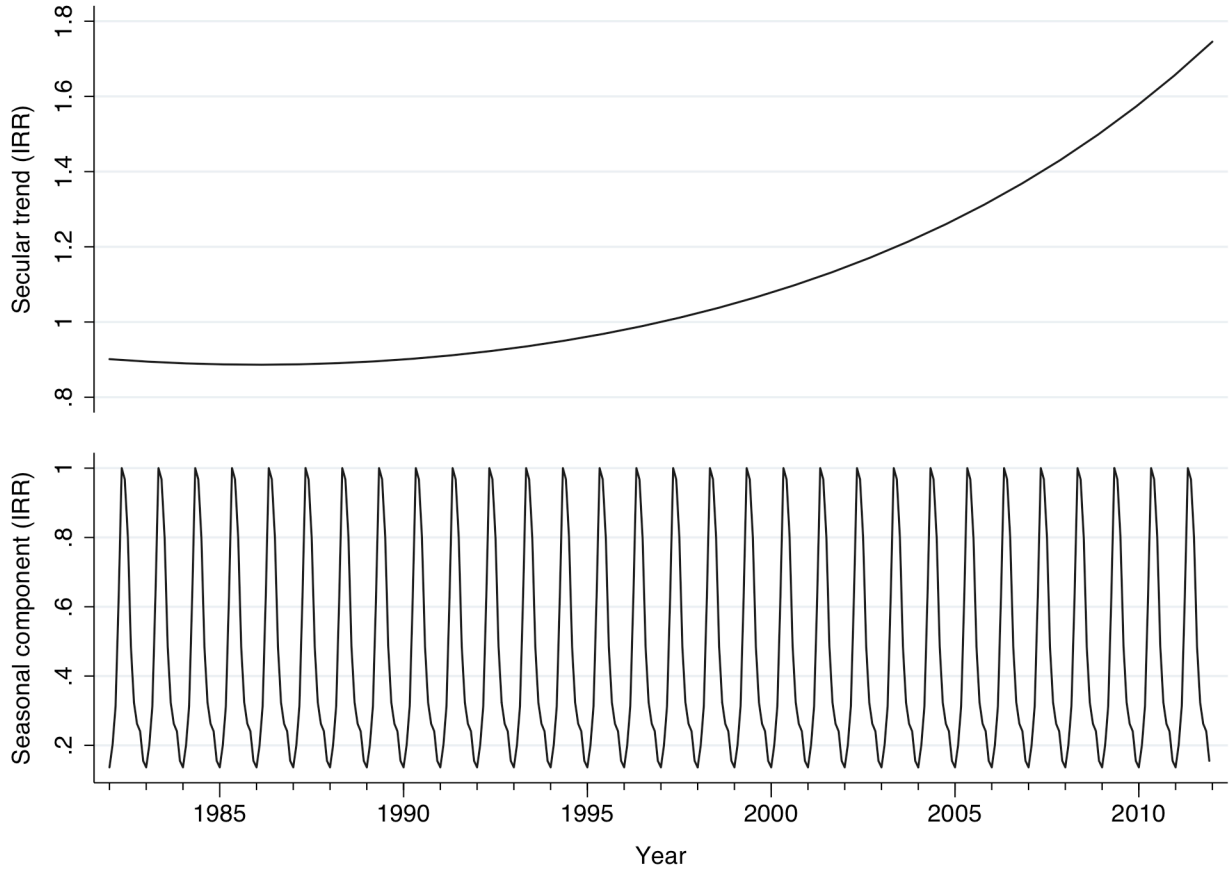

Figure 3.4: Secular trend and seasonal component of predicted hand, foot and mouth disease (HFMD) incidence, given as incidence rate ratios (IRR) relative to the center of the period (year = 1996.5) in the case of the trend component, and the peak incidence month (May) for the seasonal component.

in our model, an undetected interaction between different meteorological variables, or a threshold effect that we failed to detect. More likely, however, climate change has been just one driver, and the observed increases in HFMD stem from some combination of changing climate, land use, demographics, population dynamics, and circulating enterovirus strains.

Our model may have also understated the strength of associations between weather and HFMD due to exposure misclassification arising from the relatively coarse temporal resolution of our data. Our raw data, both weather and HFMD, were of daily temporal resolution. Our decision to work at the monthly scale, rather than weekly, was based largely on the computational demands of cokriging more than 3,000 weather surfaces (360 months, times nine weather variables) which, at more than 20 minutes computational time per surface, required over six-weeks of total running time. Developing a weekly dataset would have increased these demands by roughly four-fold.

Our data's spatial and temporal dependence precluded the use of conventional resampling methods for model validation (*e.g.* traditional cross-validation or bootstrapping). While resampling methods exist for use with correlated data, they are less well established and their properties (*e.g.* bias) are less well understood. Our analysis plan, consequently, specified split-sample validation; the data split for training and validation samples was specified *a priori*. In retrospect, however, the results of our validation—most notably, the more accurate national-level predictions in our validation than training set—suggest that our split was fortuitous and that our validation likely overestimates our model's out-of-sample predictive accuracy. Some of the larger national-level prediction errors in our training sample appear to result from the model's failure to fully capture the extent of the unusually

high peak incidence in the summer of 2010. Still, the accuracy of the national-level predictions for 2011 is uncanny, both in absolute terms and relative to the accuracy of the predictions in each of the three prior years.

The degree to which our findings may generalize outside of China, to the larger Asia-Pacific region, is not clear. China's vast area, physical adjacency to most countries in mainland Asia, and its climatic, ecological, and demographic diversity, do suggest, however, that our data may be moderately representative of the larger region. We, therefore, believe that our results may offer at least some insight into emergence and growth of HFMD in Asia and that they justify further study of the potential role of climate change in the growth of HFMD in the larger Asia-Pacific region outside of China.

We found strong associations between meteorological variables and HFMD incidence in mainland China between 2008 and 2011, and developed a weather-based predictive model for HFMD. Our model's predictions, based on historical climate data from 1982 through 2011, indicate that climate in China has changed in a manner that would favor increased HFMD, and strongly suggest that climate change has played a role in the recent growth of HFMD in China. Our results suggest that additional research is needed of the potential role of climate change in the increased incidence of HFMD, and add to the growing body of research finding potential links between climate change and human health.

SECTION C: CONCLUSIONS

These studies are applications of epidemiological and statistical methods toward three unique aims: 1) validation, 2) description, and 3) hypothesis testing. They demonstrate the application of diverse methods including analyses of spatial and time-series data; application of regression models for description, prediction and hypothesis testing; pure spatial methods including cartography and kriging; and the application of data not commonly used in epidemiological research, including climate, elevation, land cover and vegetation data. Similarly, these studies demonstrate the application of advanced data management and processing techniques, including merging disparate data, spatial data linkages, geographic data management, raster processing and the calculation of land cover pattern metrics. Beyond making for a strong technical exercise, I believe that incorporating methods and data uncommon to epidemiological research helps advance the field by broadening its scope and enhancing its toolbox.

The validation study yielded promising results: they suggest that wild-zoonoses are, in fact, more strongly associated with land cover pattern, and support the theoretical justification for using associations with land cover pattern to distinguish wild-zoonoses from other disease types. And while the wild-zoonoses were more strongly associated with land cover pattern, the nature and direction of these associations were not uniform for all three of the wild-zoonoses: the two tick-borne zoonoses (Lyme disease and Rocky Mountain Spotted Fever) had similar associations with land cover pattern metrics; however these associations were always in the opposite direction of those observed with West Nile Virus. This suggests, beyond simply identifying wild-zoonoses, the pattern of associations found between a given

disease and a set of land cover pattern metrics may be a signature of a specific reservoir, vector, or mode of transmission. It seems possible, then, that every disease could have a unique spatio-temporal signature that results from the combination of the separate spatio-temporal signatures of each of its components and their interactions; and that signature then is reflected in the disease's associations with land cover type and pattern, climate, demographics, season, and other ecological factors. With that, diseases having shared epidemiological characteristics should also have spatio-temporal signatures with common features reflecting those shared characteristics. These shared characteristics should, then, act as markers of those shared epidemiological characteristics. From this perspective, my dissertation may be the first step (among many, I'm sure) in identifying these unique signatures—and the markers of specific epidemiologic characteristics contained therein—within associations between the physical environment (*i.e.* land cover pattern and climate) and disease incidence. Clearly, more research is needed to validate these ideas, identify important signature elements and develop a functional predictive tool.

As I felt that further validation and development was first needed, I abandoned my original plans to use associations with land cover pattern to assess to potential existence of an animal reservoir for HMFD agents. Nevertheless, informally comparing the associations found with HFMD to those found with the wild-zoonoses in the validation suggests that a wild animal reservoir is very unlikely: the associations found between land cover pattern and HFMD incidence were far weaker than those found with the wild-zoonoses in the validation. This conclusion is, of course, based on an informal comparison against an incompletely validated and incompletely developed method and should be interpreted with appropriate

caution. The question remains, then, of the underlying drivers of HFMD's seasonality and its associations with meteorological variables.

Despite our incomplete understanding of these underlying drivers, HFMD's connection to weather is clear and has been found in several studies in different countries and years. Modeling the associations between meteorological variables and HFMD, and applying this model to historical climate data, allowed me to estimate the effect that climate change could have been expected to have had on HFMD incidence in mainland China during the three-decade period from 1982 through 2011. The results suggest that climate change should have caused a nearly two-fold increase in the expected incidence of HFMD in mainland China during that period; and, therefore, implicate climate change as a possible cause of the recently observed growth of HFMD in China. And while I am not the first to suggest a possible connection between climate change and HFMD emergence, I believe that the research presented herein is the first to have specifically studied the possible connection, and the first to offer clear evidence of the potential role of climate change in the increased incidence of HFMD.

These studies demonstrate the application of diverse epidemiological and statistical methods and the use of data and methods that are uncommon in epidemiologic research. The results recommend directions for future research, implicate climate change as a driver of the recent emergence of HFMD in China, offer insight into HFMD's landscape ecology, and demonstrate the potential utility of analyzing a disease's associations with land cover pattern to better understand its underlying epidemiology.

REFERENCES

1. Wong SS, Yip CC, Lau SK, Yuen KY. Human enterovirus 71 and hand, foot and mouth disease. *Epidemiol Infect.* Aug 2010;138(8):1071-1089.

2. Qiu J. Enterovirus 71 infection: a new threat to global public health? *Lancet Neurol.* Oct 2008;7(10):868-869.

3. Qiu J. Viral outbreak in China tests government efforts. *Nature.* Apr 2 2009;458(7238):554-555.

4. Higgins PB. Viruses associated with acute respiratory infections 1961-71. *J Hyg (Lond).* Jun 1974;72(3):425-432.

5. Hii YL, Rocklov J, Ng N. Short term effects of weather on hand, foot and mouth disease. *PloS one.* 2011;6(2):e16796.

6. Ma E, Lam T, Wong C, Chuang SK. Is hand, foot and mouth disease associated with meteorological parameters? *Epidemiology and infection.* Dec 2010;138(12):1779-1788.

7. Wang Y, Feng Z, Yang Y, et al. Hand, foot, and mouth disease in China: patterns of spread and transmissibility. *Epidemiology.* Nov 2011;22(6):781-792.

8. Zou XN, Zhang XZ, Wang B, Qiu YT. Etiologic and epidemiologic analysis of hand, foot, and mouth disease in Guangzhou city: a review of 4,753 cases. *Braz J Infect Dis.* Sep-Oct 2012;16(5):457-465.

9. Zhuang D, Hu W, Ren H, Ai W, Xu X. The influences of temperature on spatiotemporal trends of hand-foot-and-mouth disease in mainland China. *International journal of environmental health research.* Feb 21 2013.

10. Wang JF, Guo YS, Christakos G, et al. Hand, foot and mouth disease: spatiotemporal transmission and climate. *International journal of health geographics.* 2011;10:25.

11. Urashima M, Shindo N, Okabe N. Seasonal models of herpangina and hand-foot-mouth disease to simulate annual fluctuations in urban warming in Tokyo. *Japanese journal of infectious diseases.* Apr 2003;56(2):48-53.

12. Onozuka D, Hashizume M. The influence of temperature and humidity on the incidence of hand, foot, and mouth disease in Japan. *Sci Total Environ.* Dec 1 2011;410-411:119-125.

13. Huang Y, Deng T, Yu S, et al. Effect of meteorological variables on the incidence of hand, foot, and mouth disease in children: a time-series analysis in Guangzhou, China. *BMC Infect Dis.* Mar 13 2013;13(1):134.

14. Hu M, Li Z, Wang J, et al. Determinants of the incidence of hand, foot and mouth disease in China using geographically weighted regression models. *PloS one.* 2012;7(6):e38978.

15. Deng T, Huang Y, Yu S, et al. Spatial-temporal clusters and risk factors of hand, foot, and mouth disease at the district level in Guangdong Province, China. *PLoS One.* 2013;8(2):e56943.

16. Chen KT, Chang HL, Wang ST, Cheng YT, Yang JY. Epidemiologic features of hand-foot-mouth disease and herpangina caused by enterovirus 71 in Taiwan, 1998-2005. *Pediatrics.* Aug 2007;120(2):e244-252.

17. Chen SC, Chang HL, Yan TR, Cheng YT, Chen KT. An eight-year study of epidemiologic features of enterovirus 71 infection in Taiwan. *Am J Trop Med Hyg.* Jul 2007;77(1):188-191.
18. Ni H, Yi B, Yin J, et al. Epidemiological and etiological characteristics of hand, foot, and mouth disease in Ningbo, China, 2008-2011. *J Clin Virol.* Aug 2012;54(4):342-348.
19. Chang LY, Tsao KC, Hsia SH, et al. Transmission and clinical features of enterovirus 71 infections in household contacts in Taiwan. *JAMA.* Jan 14 2004;291(2):222-227.
20. Ruan F, Yang T, Ma H, et al. Risk factors for hand, foot, and mouth disease and herpangina and the preventive effect of hand-washing. *Pediatrics.* Apr 2011;127(4):e898-904.
21. Rajtar B, Majek M, Polanski L, Polz-Dacewicz M. Enteroviruses in water environment--a potential threat to public health. *Annals of agricultural and environmental medicine : AAEM.* Dec 2008;15(2):199-203.
22. Solomon T, Lewthwaite P, Perera D, Cardosa MJ, McMinn P, Ooi MH. Virology, epidemiology, pathogenesis, and control of enterovirus 71. *Lancet Infect Dis.* Nov 2010;10(11):778-790.
23. Zhang Y, Wang J, Guo W, et al. Emergence and transmission pathways of rapidly evolving evolutionary branch C4a strains of human enterovirus 71 in the Central Plain of China. *PLoS One.* 2011;6(11):e27895.
24. Yan XF, Gao S, Xia JF, Ye R, Yu H, Long JE. Epidemic characteristics of hand, foot, and mouth disease in Shanghai from 2009 to 2010: Enterovirus 71 subgenotype C4 as the primary causative agent and a high incidence of mixed infections with coxsackievirus A16. *Scand J Infect Dis.* Apr 2012;44(4):297-305.
25. Chang JY, Chang CP, Tsai HH, et al. Selection and characterization of vaccine strain for Enterovirus 71 vaccine development. *Vaccine.* Jan 17 2012;30(4):703-711.
26. Huang WC, Huang LM, Kao CL, et al. Seroprevalence of enterovirus 71 and no evidence of crossprotection of enterovirus 71 antibody against the other enteroviruses in kindergarten children in Taipei city. *J Microbiol Immunol Infect.* Apr 2012;45(2):96-101.
27. Li J, Chen F, Liu T, Wang L. MRI findings of neurological complications in hand-foot-mouth disease by enterovirus 71 infection. *Int J Neurosci.* Jul 2012;122(7):338-344.
28. Paticheep S, Kongsaengdao S. Viral infection of central nervous system in children: one year prospective study. *J Med Assoc Thai.* Dec 2011;94 Suppl 7:S24-31.
29. Xu W, Liu CF, Yan L, et al. Distribution of enteroviruses in hospitalized children with hand, foot and mouth disease and relationship between pathogens and nervous system complications. *Virol J.* 2012;9(1):8.
30. Breman JG, Johnson KM, van der Groen G, et al. A search for Ebola virus in animals in the Democratic Republic of the Congo and Cameroon: ecologic, virologic, and serologic surveys, 1979-1980. Ebola Virus Study Teams. *J Infect Dis.* Feb 1999;179 Suppl 1:S139-147.
31. Leroy EM, Kumulungui B, Pourrut X, et al. Fruit bats as reservoirs of Ebola virus. *Nature.* Dec 1 2005;438(7068):575-576.
32. Aron JL, Patz J. *Ecosystem change and public health : a global perspective.* Baltimore: Johns Hopkins University Press; 2001.

33. Kilpatrick AM, Altizer S. Disease ecology. *Nature Education Knowledge.* 2010;2(12):13.
34. Levin SA, Carpenter SR. *The Princeton guide to ecology.* Princeton: Princeton University Press; 2009.
35. Chen L, Mou X, Zhang Q, et al. Detection of human enterovirus 71 and coxsackievirus A16 in children with hand, foot and mouth disease in China. *Mol Med Report.* Apr 2012;5(4):1001-1004.
36. Li Y, Zhu R, Qian Y, Deng J. The characteristics of blood glucose and WBC counts in peripheral blood of cases of hand foot and mouth disease in China: a systematic review. *PLoS One.* 2012;7(1):e29003.
37. Li R, Zou Q, Chen L, Zhang H, Wang Y. Molecular analysis of virulent determinants of enterovirus 71. *PLoS One.* 2011;6(10):e26237.
38. Hagiwara A, Tagaya I, Komatsu T. Seroepidemiology of enterovirus 71 among healthy children near Tokyo. *Microbiol Immunol.* 1979;23(2):121-124.
39. Yin W. Fighting against infectious diseases in China. *Hum Vaccin.* Dec 2011;7(12):1250-1253.
40. Chumakov M, Voroshilova M, Shindarov L, et al. Enterovirus 71 isolated from cases of epidemic poliomyelitis-like disease in Bulgaria. *Arch Virol.* 1979;60(3-4):329-340.
41. Hagiwara A, Tagaya I, Yoneyama T. Epidemic of hand, foot and mouth disease associated with enterovirus 71 infection. *Intervirology.* 1978;9(1):60-63.
42. Wakefield J. Sensitivity analyses for ecological regression. *Biometrics.* Mar 2003;59(1):9-17.
43. Jourdain E, Gauthier-Clerc M, Bicout DJ, Sabatier P. Bird migration routes and risk for pathogen dispersion into western Mediterranean wetlands. *Emerg Infect Dis.* Mar 2007;13(3):365-372.
44. Rappole JH, Derrickson SR, Hubalek Z. Migratory birds and spread of West Nile virus in the Western Hemisphere. *Emerg Infect Dis.* Jul-Aug 2000;6(4):319-328.
45. Reed KD, Meece JK, Henkel JS, Shukla SK. Birds, migration and emerging zoonoses: west nile virus, lyme disease, influenza A and enteropathogens. *Clin Med Res.* Jan 2003;1(1):5-12.
46. MacLennan C, Solomon T. Potential neurovirulence of common cold virus. *Lancet.* Nov 20-26 2004;364(9448):1839-1840.
47. Leitão AB. *Measuring landscapes : a planner's handbook.* Washington, DC: Island Press; 2006.
48. McGarigal K, Cushman SA, Ene E. FRAGSTATS v4: Spatial Pattern Analysis Program for Categorical and Continuous Maps. 2012; http://www.umass.edu/landeco/research/fragstats/documents/fragstats.help.4.pdf. Accessed December 3, 2012.
49. Holben BN. Characteristics of maximum-value composite images from temporal AVHRR data. *International Journal of Remote Sensing.* 1986;7(11):1417-1434.
50. Bontemps S, Defourny P, Van Bogaert E, Arino O, Kalogirou V, Perez JR. *GlobCover 2009: Products description and validation report.* European Space Agency;2011.
51. Tucker CJ, Pinzon JE, Brown ME, et al. An extended AVHRR 8-km NDVI dataset compativle with MODIS and SPOT vegetation NDVI data. *International Journal of Remote Sensing.* 2005;26(20):4485-4498.

52. USGS. *Shuttle Radar Topography Mission Digital Elevation Model.* Global Land Cover Facility, University of Maryland, College Park, Maryland2004.
53. Qiaoyun F, Xiongfei J, Lihuan L, Angao X. Epidemiology and etiological characteristics of hand, foot and mouth disease in Huizhou City between 2008 and 2011. *Arch Virol.* Apr 2013;158(4):895-899.
54. Zhu Q, Hao Y, Ma J, Yu S, Wang Y. Surveillance of hand, foot, and mouth disease in mainland China (2008-2009). *Biomed Environ Sci.* Aug 2011;24(4):349-356.
55. Turner MG, Gardner RH, Dale VH, O'Neill RV. Predicting the spread of disturbance across heterogeneous landscapes. *Oikos.* 1989;55(1):121-129.
56. Turner MG. Landscape ecology: the effect of pattern on process. *Annual Review of Ecology and Systematics.* 1989;20:171-197.
57. NOAA. *Global Surface Summary of the Day - GSOD.* Asheville, NC2011.
58. Burnham KP, Anderson DR, Burnham KP. *Model selection and multimodel inference : a practical information-theoretic approach.* 2nd ed. New York: Springer; 2002.