Geosciences **2014**, *4*, 202-218; doi:10.3390/geosciences4030202

OPEN ACCESS

geosciences

ISSN 2076-3263

www.mdpi.com/journal/geosciences

Article

Risk Factors for *E. coli* O157 and Cryptosporidiosis Infection in Individuals in the Karst Valleys of East Tennessee, USA

Ingrid Luffman [1,*] and Liem Tran [2]

[1] Department of Geosciences, East Tennessee State University, Johnson City, TN 37614-1709, USA

[2] Department of Geography, The University of Tennessee, Knoxville, TN 37996-0925, USA; E-Mail: ltran1@utk.edu

* Author to whom correspondence should be addressed; E-Mail: luffman@etsu.edu; Tel.: +1-423-439-7551; Fax: +1-423-439-7520.

Received: 25 June 2014; in revised form: 14 August 2014 / Accepted: 15 August 2014 / Published: 27 August 2014

Abstract: This research examines risk factors for sporadic cryptosporidiosis and *Escherichia coli* (*E. coli*) O157 infection in East Tennessee, using a case-control approach and spatial logistic regression models. The risk factors examined are animal density, land use, geology, surface water impairment, poverty rate and availability of private water supply. Proximity to karst geology, beef cow population density and a high percentage of both developed land and pasture land are positively associated with both diseases. The availability of private water supply is negatively associated with both diseases. Risk maps generated using the model coefficients show areas of elevated risk to identify the communities where background risk is highest, so that limited public health resources can be targeted to the risk factors and communities most at risk. These results can be used as the framework upon which to develop a comprehensive epidemiological study that focuses on risk factors important at the individual level.

Keywords: environmental risk factors; spatial modeling; cryptosporidiosis; *E. coli* O157; spatial epidemiology; spatial logistic regression

1. Introduction

The field of medical geology assesses health problems associated with geologic materials with three areas of focus: (1) geology as a source of harmful materials; (2) movement and alteration of harmful

materials through the subsurface over time and space; and (3) exposure pathways associated with geologic materials [1]. This research focuses on the third branch, specifically on karst geology and other spatially-distributed risk factors, as pathways for exposure to waterborne diseases.

The aim for epidemiologic research in general is to identify associations between exposures and outcomes to maximize health or to prevent disease [2], and the probability of human infection by pathogens depends on a number of factors, including how well the pathogen survives in the environment and the opportunities for host-pathogen interaction [3].

It is well established that karst regions are at a higher risk for groundwater contamination due to groundwater-surface water interactions and low groundwater residence times [4,5]. Natural and anthropogenic processes impact water quality at karst springs, as surface contamination is quickly carried into the groundwater supply when contaminated surface runoff flows into sinkholes and sinking streams [6,7]. Contaminated groundwater supplies used for public or private water supplies can result in outbreaks of disease that are more prevalent in karst regions [8,9].

Exposure to impaired surface water [10–13], agricultural activity [10,13–16] and karst geology [17] has been linked to outbreaks of cryptosporidiosis and *Escherichia coli* O157 (*E. coli* O157) infection worldwide. Since the first outbreak of cryptosporidiosis related to recreational water in the United States was reported in 1988, *Cryptosporidium* has emerged as the most recognized cause of disease outbreak associated with recreational water [18], as it is pervasive in the environment, resistant to chlorine and has a low infectious dose (10 to 30 oocysts) [19]. From 2006 to 2009, Tennessee reported 315 cryptosporidiosis cases statewide. Forty-nine (15.7%) were from the northeast region, though this region represents only 6% of the state's population. Forty-seven of the forty-nine cases (96%) in the northeast region were in two counties, and while some of these cases were attributed to a specific exposure (such as contaminated food or water), the sources of infection for most of the cases remained unexplained.

E. coli O157 is a pathogen first identified in 1982 as the cause of two outbreaks of disease in Oregon and Michigan, USA [18]. Since that time, the disease has become widely distributed throughout the United States and the rest of the world because of the high survival rate of the pathogen and low infectious dose (between 10 and 100 organisms) [11]. *E. coli* O157 infection is associated with consumption of contaminated water or food, such as undercooked beef, dairy products and salads; however, a connection between environmental exposure and *E. coli* O157 outbreaks has also been established [11,14,18,20,21]. *E. coli* O157 infection causes an estimated 96,534 illnesses in the United States each year, 3268 of which can require hospitalization [22]. From 2000 to 2010, 903 cases of *E. coli* O157 infection were reported in Tennessee. In 21 cases, the onset of symptoms was preceded by international travel, and only two cases were associated with a known outbreak. Therefore, in the majority of the Tennessee cases, the cause of illness is unknown.

Because known risk factors for cryptosporidiosis and *E. coli* O157 infection are associated with an individual's environment (apart from exposure through food and human contact), an analysis of these datasets would benefit from explicitly including space and spatial relationships between potential risk factors and disease. Proximity to a known risk factor may increase the incidence of cryptosporidiosis or *E. coli* O157 infection in a population, and therefore, epidemiologic research should take into account the spatial relationship between the individual, the environment and other individuals, keeping in mind the relationship between and among cases of these diseases. A geographic approach to the assessment of risk for cryptosporidiosis and *E. coli* O157 infection that includes the use of GIS and

spatial statistical modeling can be a powerful method to infer associations between the environment and health [23].

Much of East Tennessee falls within the Valley and Ridge physiographic province of North America, characterized by folded Paleozoic sedimentary rocks (limestone, shale and sandstone) with flat-lying sedimentary rocks to the west and Precambrian metamorphic rocks of the Blue Ridge province to the east [24]. Building on the established link between karst geology, water quality and health, this research examines the role of karst geology and other environmental risk factors for cryptosporidiosis and *E. coli* O157 infection in East Tennessee.

2. Methods

The research was accomplished in two steps. First, spatial databases of disease data and explanatory variables were assembled, and exploratory mapping was done. In this step, the cases were geocoded, rates were calculated and standardized for each zip code and explanatory variables were extracted for each case and zip code and overlaid with the disease data. In the second step, regression models were developed using a case-control approach to examine the risk for disease in the individual.

2.1. Data

For this study, a dataset of patient records for 903 *E. coli* O157 infection and 555 cryptosporidiosis cases occurring in Tennessee from 2000 to 2010 was extracted from the Foodborne Diseases Active Surveillance Network (FoodNet) database, USA Department of Health and Human Services. The patient records consisted of patient age, gender, date of specimen, address, recent travel history and whether the infection was associated with an outbreak. Appropriate safeguards were taken to protect patient privacy and confidentiality. Case records for East Tennessee addresses were extracted and examined for duplicates, recent travel, association with a known outbreak and incomplete records, and geocoding was completed using the University of South California WebGIS Services batch geocoder [25]. After cleaning, the datasets contained 247 cryptosporidiosis and 250 *E. coli* O157 infection records (Figure 1).

The environmental risk factors selected as explanatory variables were geology, surface water impairment, agricultural animal population density by zip code, land use, groundwater well permit density by county and poverty rate at the block group level (Table 1). Each explanatory variable was extracted at the best resolution available. These data are publicly available as spatially-referenced spreadsheet data (related to a geographic unit such as a county or census unit) or as shapefiles that were directly imported into ArcGIS 10.0 [26].

Karst regions were delineated by selecting polygons classified as limestone or dolomite in the 1:250,000 Geology of Tennessee shapefile [27] and creating a new layer of the karst-prone regions of Tennessee. Surface water shapefiles and attribute data for the Year 2008 305(b) Report [28] and Year 2008 303(d) Report [29] were downloaded from the Environmental Protection Agency Reach Address Database [30]. Using the near tool in ArcGIS, two raster files containing values for the distance in kilometers to the nearest karst area (KARST) and distance to the nearest impaired stream segment (STREAM) were constructed.

Agricultural animal population densities were calculated for each zip code tabulation area (ZCTA) using the USDA Agricultural Census Data from 2007 [31] for dairy cows (MILK), beef cows (BEEF),

hogs (HOG) and sheep (SHEEP). The 2006 National Land Cover Dataset [32] was reclassified and smoothed to construct three 500 m × 500 m rasters of percent cover for each of three land use classes (PASTURE, FOREST and DEVELOPED). A database of well permits was obtained from the Tennessee Department of Environment and Conservation, summed by county and joined to the 2010 Tiger/Line county layer (WELLS). Poverty rates at the block group level (POVRATE) were obtained from USA Census data [33]. Raster layers for each variable were constructed in ArcGIS for input as spatial explanatory variables in the spatial logistic regression model (Figure 2).

Figure 1. Study area and approximate locations of disease cases from 2000 to 2010.

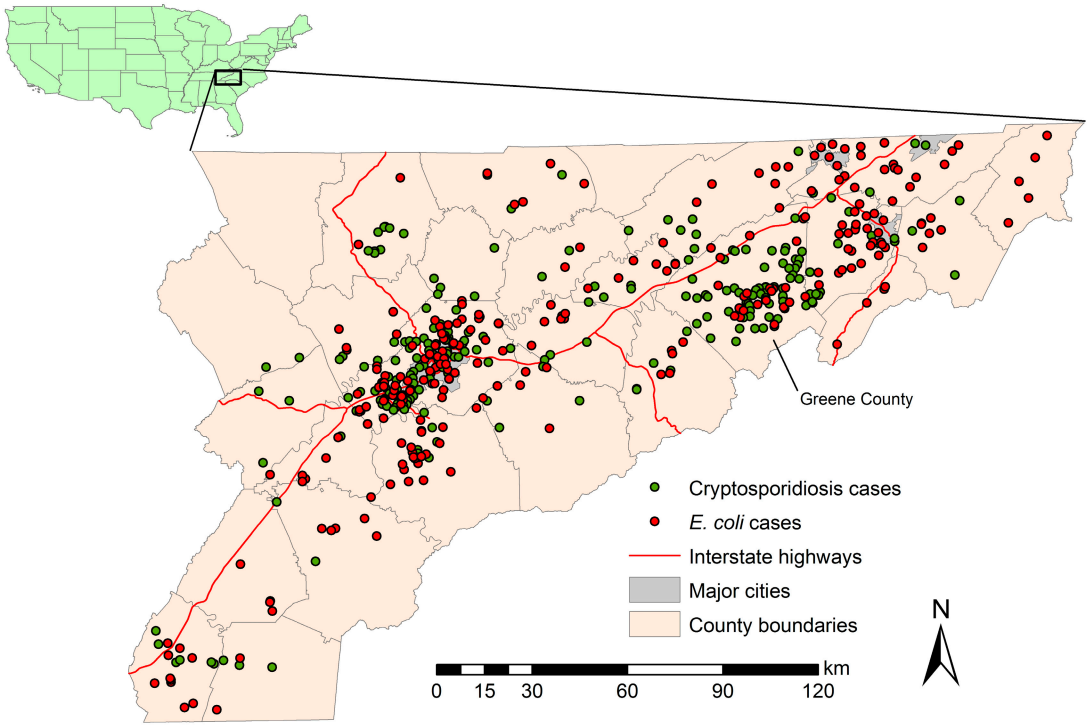
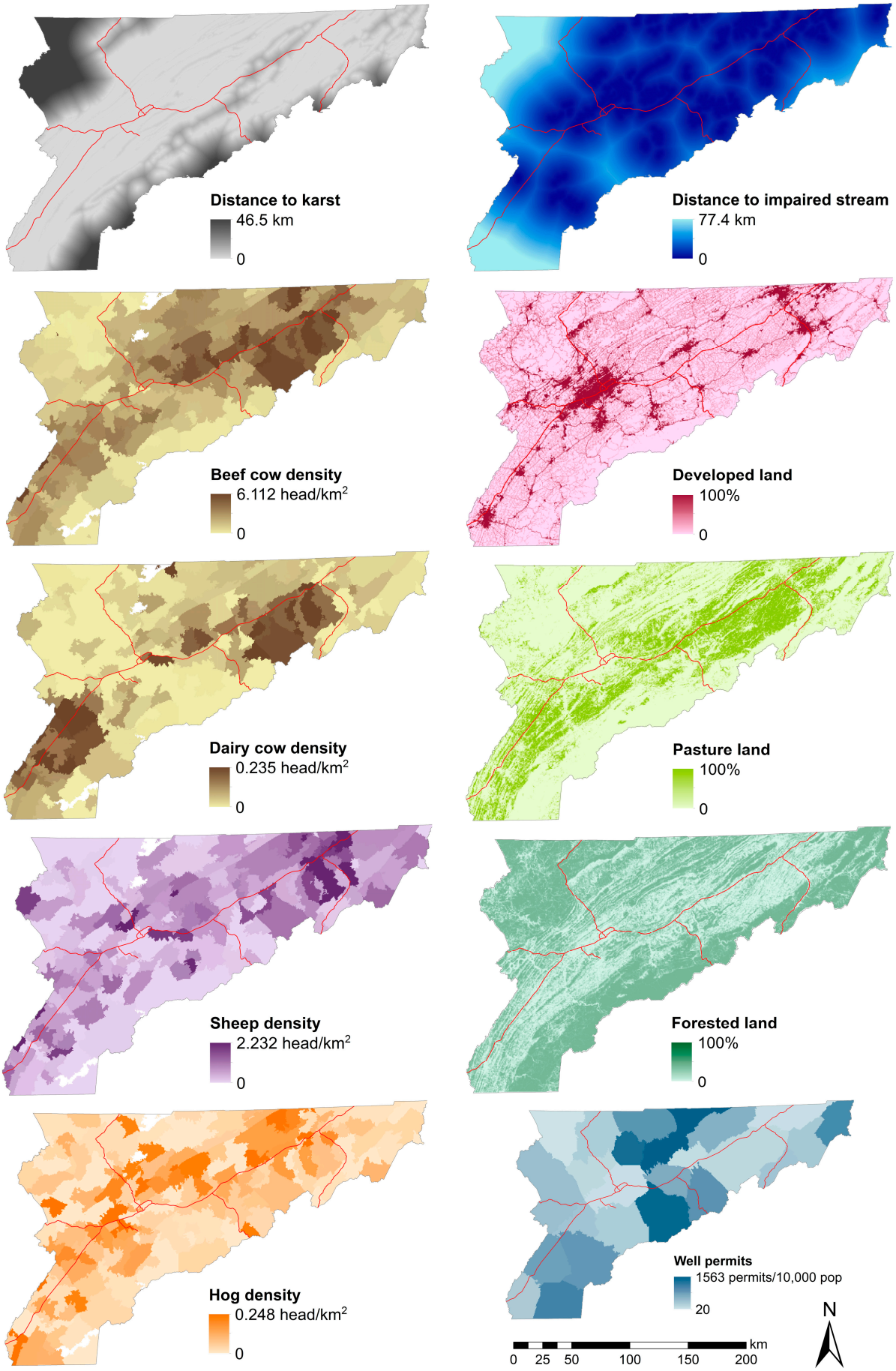


Table 1. Candidate explanatory variables for cryptosporidiosis and *E. coli* O157 infection.

Variable	Description	Units
KARST [1]	Distance to nearest karst geology	Kilometers (Euclidean distance)
STREAM [2]	Distance to nearest *E. coli* contaminated stream segment	Kilometers (Euclidean distance)
BEEF [3]	Beef cow population density (zip code level data)	Animals/km^2
MILK [3]	Dairy cow population density (zip code level data)	Animals/km^2
HOG [3]	Hog population density (zip code level data)	Animals/km^2
SHEEP [3]	Sheep population density (zip code level data)	Animals/km^2
DEVELOPED [4]	Percent cover in a 500 m × 500 m raster cell	Percent, expressed as decimal
FOREST [4]	Percent cover in a 500 m × 500 m raster cell	Percent, expressed as decimal
PASTURE [4]	Percent cover in a 500 m × 500 m raster cell	Percent, expressed as decimal
WELLS [5]	Number of well permits by population (county level data)	Permits/10,000 population
POVRATE [6]	Poverty rate (block group level data)	Percent, expressed as decimal

Data Sources: [1] [27]; [2] [28,29];[3] [30]; [4] [31]; [5] TDEC Well permit database; [6] [33].

Figure 2. Environmental risk factors used as covariates in the spatial logistic regression model.



2.2. Spatial Logistic Regression Model

The objective of a case-control modeling approach is to identify and quantify the relationship between risk factors and cryptosporidiosis and *E. coli* O157 infection in the individual. One limitation of a case-control study using an existing database is that control data may not be available, in which case random absence data can be generated and used as the control dataset [34]. To avoid this issue, a spatial logistic regression model was developed in R [35], using the `slrm` function (R package `spatstat`) [36]. The case data were entered as a point process, and the covariate data were input as image files that span the area of interest.

The explanatory variables selected for inclusion in cryptosporidiosis models were: beef and dairy cow population density; sheep population density; percent of developed, forested and agricultural land; distance to karst geology; distance to impaired stream; poverty rate; and well permit density. Hog population density was not included in the model, because exploratory analyses indicated there was no significant statistical correlation between hog population density and cryptosporidiosis.

Explanatory variables selected for inclusion in the *E. coli* O157 models were: beef and dairy cow population density; hog population density; percent of developed, forested and agricultural land; distance to karst geology; distance to impaired stream; poverty rate; and well permit density. Sheep population density was not included in the model, because there was no significant statistical correlation between sheep density and *E. coli* O157 infection in preliminary analyses.

Risk maps were generated in ArcGIS 10.0 for both cryptosporidiosis and *E. coli* O157 infection from the model coefficients using the raster calculator tool and the equation:

$$P = [1 + \exp(-\log a - BX)]^{-1} \quad (1)$$

where P is the risk for disease at any given raster cell, $\log a$ is an offset term representing log-transformed population, B is the vector of model coefficients and X is the vector of covariate values. The Akaike information criterion (AIC), a diagnostic tool to quantify the trade-off between increased explanatory power and information loss associated with the use of additional explanatory variables in a model, was used to select the best model for each disease.

3. Results

3.1. Cryptosporidiosis

Multiple spatial logistic regression models were generated in R using the `slrm` function, with different combinations of covariates and interaction terms (Table 2). Model C1 includes all candidate covariates, except HOG, which showed no relationship with cryptosporidiosis. Model C2 includes only those variables significant in Model C1, and C3 includes all variables significant in C1 plus all possible interaction terms. Interactions were examined, but were not statistically significant ($p > 0.05$). Model results were generally consistent between models, and Model C1 was the best model, using the Akaike information criterion (AIC) as a diagnostic tool. Given the order of the AIC, the range of 8957–8968 between the three models is not large, indicating minimal additional loss of information with increased model complexity.

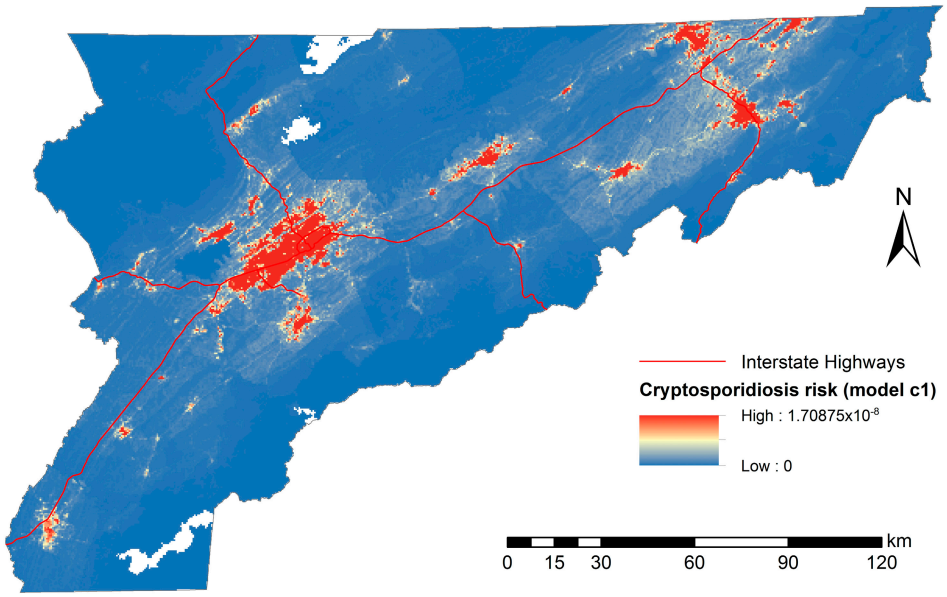
Table 2. Spatial logistic regression model (SLRM) coefficients (*p*-values in parentheses) for cryptosporidiosis individual models. Shaded cells represent excluded variables. Model diagnostics (AIC) are shown on the bottom row.

Covariates	SLRM Coefficients (*p*-values)		
	C1	C2	C3
(Intercept)	−18.58868	−19.28357	−19.54150
BEEF	0.1899209 (0.000)	0.415818 (0.000)	0.384648 (0.000)
MILK	9.700272 (0.599)		
SHEEP	−6.439952 (0.973)		
DEVELOPED	0.02983753 (0.000)	0.034451 (0.000)	0.041139 (0.000)
PASTURE	0.01080318 (0.000)	0.016177 (0.000)	0.012557 (0.000)
FOREST	−0.0055259 (0.761)		
WELLS	−0.001641 (0.000)	−0.001450 (0.000)	−0.000690 (0.000)
KARST	−0.0001505 (0.000)	−0.0001545 (0.000)	−0.000518 (0.000)
STREAM	−0.0000237 (0.004)	−0.0000198 (0.014)	−0.0000029 (0.014)
POVRATE	−0.3994621 (0.538)		
Model Diagnostics			
AIC	8957.298	8961.791	8968.142

The factors most significant for increased risk for cryptosporidiosis are proximity to karst geology (KARST), proximity to *E. coli* impaired streams (STREAM), higher beef cow population density (BEEF), residence within developed (DEVELOPED) or agricultural (PASTURE) land and a lower well density by population (WELLS).

A map showing background environmental risk generated using Model C1 (Figure 3) and Equation (1) using an offset of log(population density) [37] shows elevated risk (as the probability of disease) along developed corridors, in urban centers and in one predominantly rural area (Greene County area). Faint northeast-southwest trending linear areas of high risk show the contribution to risk from karst geology.

Figure 3. Spatial logistic regression model (C1) risk map for cryptosporidiosis (see Table 2 for the coefficients).



3.2. E. coli O157

Spatial logistic regression models for *E. coli* O157 infection were developed in R for various combinations of covariates and interaction terms (Table 3). Results were generally consistent between models. Model E1 included all candidate variables, with no interaction terms. Model E2 included only the significant variables from E1, and a comparison of AIC shows that E2 is a slight improvement over E1. Model E3 was constructed using only the variables in E2 plus all possible interaction terms. Only two interaction terms (DEVELOPED × PASTURE and HOG × KARST) emerged as significant, and the AIC increased due to the increased complexity in the model.

Table 3. Spatial logistic regression model coefficients (*p*-values in parentheses) for *E. coli* O157 individual models. Shaded cells represent excluded variables. Model diagnostics (AIC) are shown in the bottom row.

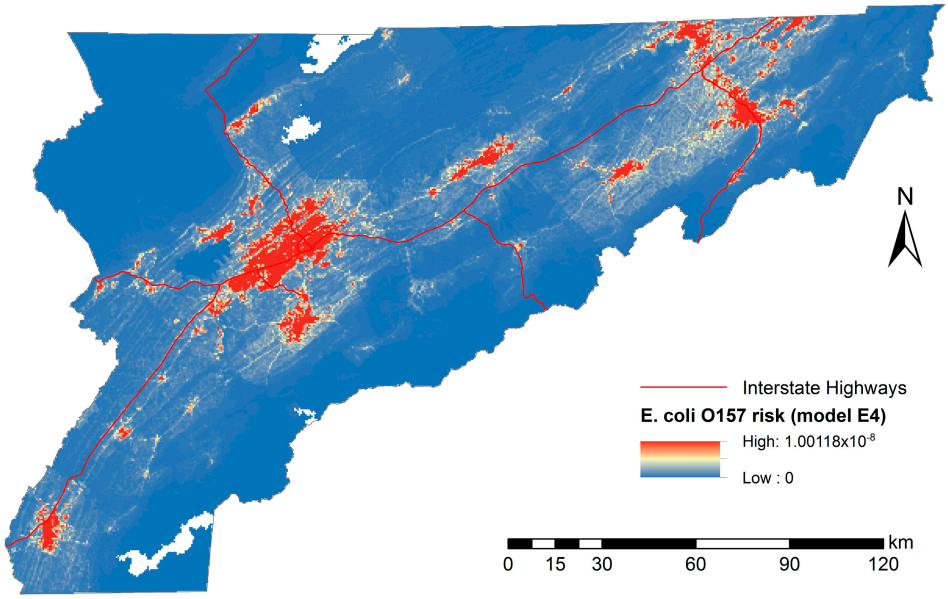
Covariates	**SLRM Coefficient (*p*-values)**			
	E1	**E2**	**E3** †	**E4**
Constant	−19.43082	−19.27842	−19.56200	−19.14675
BEEF	0.3765216 (0.000)	0.3863660 (0.000)	1.633936 (0.000)	0.3046392 (0.000)
MILK	−2.96153 (0.001)	−0.299347 (0.001)	−20.07985 (0.001)	−1.319051 (0.001)
HOG	3.016985 (0.023)	2.954141 (0.023)	4.032882 (0.023)	1.794563 (0.023)
DEVELOPED	0.0322240 (0.000)	0.0307588 (0.000)	0.0320371 (0.000)	0.0282152 (0.000)
PASTURE	0.0166063 (0.000)	0.0150557 (0.000)	0.0093708 (0.000)	0.0103302 (0.000)
FOREST	0.0187083 (0.737)			
WELLS	−0.00122612 (0.000)	−0.00122976 (0.000)	−0.000922645 (0.000)	−0.00128114 (0.000)
KARST	−0.000219501 (0.000)	−0.000216085 (0.000)	−0.000380838 (0.000)	−0.000336388 (0.000)
STREAM	0.0000023 (0.721)			
POVRATE	0.184116 (0.774)			
HOG*KARST			0.00698929 (0.036)	0.00518194 (0.064)
DEVELOPED*PASTURE			0.000405056 (0.005)	0.000354069 (0.003)
Model Diagnostics				
AIC	9190.348	9184.642	9194.744	9176.378

† Model E3 included all possible interactions between explanatory variables, but only coefficients for significant interactions are presented in the table to conserve space. Model E4 included only the two interaction terms listed.

The final model (E4) was the best SLRM (using the AIC as a diagnostic tool) and included only significant covariates from previous models. The risk factors for *E. coli* O157 infection identified in Model E4 are: distance to karst geology (KARST); beef cow (BEEF), dairy cow (MILK) and hog (HOG) population density; percent of developed land (DEVELOPED); percent of agricultural land (PASTURE); and well permit density (WELLS). The interaction term DEVELOPED*PASTURE was significant ($p < 0.05$). Disease is positively associated with increasing values of BEEF, HOG, DEVELOPED and PASTURE and negatively associated with increasing values of MILK, WELLS and KARST.

A map of background environmental risk generated using Model E4 (Figure 4) and Equation (1) with an offset of log(population density) shows elevated risk along developed corridors and in urban centers, in areas of agricultural land use and in areas of karst.

Figure 4. Spatial logistic regression model (E4) risk map for *E. coli* O157 infection (see Table 3 for the coefficients).



4. Discussion

Both cryptosporidiosis and *E. coli* O157 infection are waterborne diseases with low infectious doses and high survivability in the environment. It is reasonable therefore to see similar environmental risk factors for both diseases. However, there are also notable differences, which will be discussed in the sections that follow.

4.1. Karst Geology

Proximity to karst geology was a small, but consistently significant, risk factor for both diseases. This finding is in agreement with previous research for cryptosporidiosis [17] and is not surprising for a number of reasons. First, karst areas are underlain by limestone that is less resistant to weathering and erosion. Karst areas therefore coincide with valley bottoms in East Tennessee, which is where development is concentrated due to the ease of construction and access to transportation corridors (roads and rivers). This may lead to increased opportunities for human-to-human transmission. Second, and perhaps more importantly, in karst areas, surface water and groundwater are well connected through sinkholes and springs, and so, surface contamination readily enters the groundwater system and can emerge elsewhere as a spring, having had little opportunity for the natural attenuation typical of non-karst regions. Patients may unwittingly become infected through contact with an impaired spring. Proximity to karst geology is therefore an important risk factor and should be included in future studies of environmental risk factors for cryptosporidiosis and *E. coli* O157 infection.

4.2. Surface Water Quality

For cryptosporidiosis, proximity to an *E. coli*-impaired stream (STREAM) was associated with increased risk for disease, which is in agreement with risk factors reported in other studies [38–40]. Contact with contaminated water may be more likely the closer the patient resides to the stream,

although it is important to note that proximity to a stream does not necessarily mean that contact occurs between an individual and the contaminant, and no data were available to determine whether individuals with either disease had physical contact with recreational water. Nevertheless, in East Tennessee, the general pattern of land development is for forested lands at higher elevations along ridge tops and for developed lands to hug the valley bottoms near established streams. The fluviokarst hydrology, characterized by both surface and subsurface drainage, sinkholes and abundant springs, provides ample opportunity for recreational contact.

Public water suppliers in the study area withdraw 82.2% of water by volume from surface lakes or streams, 15.9% from wells and 10.5% from springs [41]. Moreover, in karst areas with fluviokarst hydrology, groundwater-surface water interactions are well established, and surface contamination may quickly enter groundwater supplies and contaminate wells and springs. Spring response to precipitation includes increased flow and suspended sediment load [42], further evidence for groundwater-surface water interactions.

For *E. coli* O157 infection, proximity to an *E. coli*-impaired stream was not an important risk factor for disease, which contrasts previous studies that associated recreational water exposure with increased risk for *E. coli* O157 infection. [13,14]. A study of 350 *E. coli* O157 outbreaks in the United States from 1982 to 2002 determined that in 9.5% of the cases, the transmission route was related to recreational water contact, 4.5% was related to contaminated water supply, 3% was attributed to cattle contact and 21% was of unknown transmission (the remainder of cases were attributed to foodborne or person-to-person transmission) [10]. Surface water streams that are recognized by the state and by the Environmental Protection Agency as impaired due to the presence of *E. coli* are considered a public health threat, because *E. coli* is an indicator organism for the presence of both cryptosporidiosis and *E. coli* O157 in surface water. Environmental contamination of surface water supplies, resulting in the presence of fecal bacteria in streams and lakes, is known to increase the risk for bacteria-related sickness and death in humans [2,18,38,42]. Nevertheless, research that employs surface water quality spatial databases in spatial disease models is not well represented in the literature. The spatial models developed in this research associate proximity to an *E. coli*-impaired surface water stream with cryptosporidiosis, but not with *E. coli* O157 infection. While *E. coli* is the standard indicator for fecal pollution used by the state of Tennessee for the presence of both *E. coli* O157 and *Cryptosporidium* in surface water, prior research has shown that it is a poor indicator for the presence of *E. coli* O157 (in one study, the pathogen was detected in less than 1% of samples testing positive for *E. coli* [43]), but it is a reasonable indicator for *Cryptosporidium* oocysts (in the same study, oocysts were detected in 40% of samples testing positive for *E. coli*). Similarly, *E. coli* and turbidity were positively correlated in surface water samples, but *E. coli* O157 was not significantly correlated with turbidity [44]. The findings of the research presented here support the questioning of the use of *E. coli* as an appropriate indicator bacteria for *E. coli* O157 and as an indicator for the public health threat.

4.3. Land Use

The percent of agricultural land (PASTURE) and the percent of developed land (DEVELOPED) were associated with increased risk for both diseases, suggesting multiple pathways of transmission. A high percentage of developed land can be thought of as a proxy for high population density, and a

higher number of cases of disease would be expected in these areas, due to the number of residents and consequent opportunities for human-to-human transmission. Agricultural land was also associated with increased disease in the individual and in the population, likely due to the presence of agricultural animals and opportunities for animal-to-human transmission [40,45].

The data used in this study were not sufficiently detailed to speciate the cryptosporidiosis cases into *C. hominis* and *C. parvum*, which are the two species commonly associated with human-to-human contact and animal-to-human contact, respectively. Others have found a positive association between developed land and *C. hominis* and between agricultural land and with *C. parvum* [46], but there is no way to test this using the cryptosporidiosis case data available for this study.

The association of developed land with *E. coli* O157 infection contrasts two other studies that positively associated disease with rural areas [14] and [47] farm density. Their findings agree, however, with the positive association between disease and agricultural land found in this study. The significance of contrasting land use variables as risk factors for both diseases suggests multiple pathways for infection. In developed land, transmission may be by human-to-human contact, foodborne transmission or through direct animal contact during farm visits. In pasture and forested lands, transmission may be through environmental contamination, foodborne transmission or through animal-to-human contact. No data about the patient's occupation was available in the dataset; however, it is reasonable to recognize that farmers, farm families and farm workers will have different exposures and immunity to both diseases, which may influence model results. The significance of the interaction term between developed land and agricultural land for *E. coli* O157 infection can be thought of as a variable expressing the proximity of development to agriculture (*i.e.*, locations where developed land and agricultural land are both dominant), providing increased opportunity for contact between residents and agricultural activity, which has been related to increased risk for *E. coli* O157 infection [13,14].

4.4. Agricultural Animal Population Densities

The positive association of beef cow population density (BEEF) with increased risk for disease agrees with other studies that have linked cattle populations with cryptosporidiosis [46,48] and *E. coli* O157 infection [14,20,21]. Although beef cow population density was positively associated with both diseases, dairy cow population density was not significantly associated with cryptosporidiosis and was negatively associated with *E. coli* O157 infection. Because of the differences in how their populations are associated with the two diseases, it is therefore important to differentiate between beef and dairy cattle in similar studies. All cattle-related studies reviewed for this research treated cattle as one group, but this research suggests that the cattle population should be partitioned into beef and dairy. The protective effect of dairy cow population density may result from the regulation of farming practices in dairy farms, specifically the controls required to capture and treat runoff, which reduce the pathogen load in the environment. In contrast, beef cattle are more common in East Tennessee, and no environmental regulations exist to control runoff from pasture land nor to control access to streams by cattle. Consequently, runoff from agricultural fields is a non-point source of surface water pollution, which increases the environmental pollution load. Interpreting the negative association between dairy cattle and *E. coli* O157 infection as a protective effect is a plausible hypothesis, but cannot be confirmed at this point.

The positive association found between hog population density (HOG) and *E. coli* O157 infection contrasts with two other studies that found a negative association [15] and no association [21].

4.5. Poverty Rate

Poverty rate was not retained as an explanatory variable for either disease, which may indicate that poverty rate at the block group level is not a good indicator of risk for disease in the individual. A better variable to include in the individual model may be the socioeconomic status of the individual, rather than the individual's block group, because the aggregate variable smoothes out variability, and there is no way to know the actual socioeconomic status of the individual when using aggregate data.

4.6. Well Permits

Well permit density was positively related to cryptosporidiosis and *E. coli* O157 infection. While not indicative of drinking water source for a given case of disease, well permit density can indicate whether private water supply is more or less common in a county, so it is a measure of the availability of a private water supply. The finding that increased disease risk in the individual is associated with decreased well density indicates that the availability of private water supplies has a protective effect. Wells in the study area have an average depth of 309 feet, which easily bottom within the karst limestone bedrock unit that underlies much of the study area. In fact, one eighth of the deepest caves in the USA are found within karst areas of Tennessee, and because these caves regularly descend beyond the average depth of the private wells, the depth of the wells is not a compelling explanation for their apparent protective effect. Note, however, that this finding agrees with a recent Tennessee Department of Health clinical epidemiological study on cryptosporidiosis conducted in the Greene County, Tennessee, region from 2009 to 2011. This study used a supplemental form to gather drinking water source, recreational water exposure, food, travel and animal exposures at the individual level. No association with private water supplies was found. This result, however, contrasts with other studies that associated private water supply with increased risk for *E. coli* O157 infection [14] and cryptosporidiosis [17,48].

4.7. Limitations

This study examines environmental risk factors for two diseases that can also be transmitted through other exposures related to an individual's behavior or lifestyle. These variables may include consumption of undercooked beef or unpasteurized dairy products, visits to farms, contact with infected persons and playing or swimming in contaminated surface water [3,11,14,45,46,49–51]. The individual models developed here do not capture this individual behavior, but instead, can capture potential exposure to risk for the individual by inclusion of environmental variables at the local scale. Proximity to contaminated surface water or to agricultural activity can increase the likelihood that a person will come in contact with an environmental reservoir of cryptosporidiosis or *E. coli* O157 and in this way can increase background risk. The individual model, therefore, can be thought of as the environmental background risk at the individual level (large-scale environmental risk).

Selection of a "best" model is secondary to identifying risk factors for *E. coli* O157 infection and cryptosporidiosis from a set of potential environmental and socioeconomic explanatory variables. An understanding of the environmental and socioeconomic variables that emerge as significant risk factors for disease can help to inform policies to combat disease prevalence. Often, these relationships can be discerned by exploratory mapping of disease cases overlaid with risk factors; however, the modeling process is important to statistically quantify the risk and the important risk factors. Risk maps, such as those displayed in Figures 2 and 3, can visually communicate the environmental background risk for disease within the study area. Because cryptosporidiosis and *E. coli* O157 infection are rare diseases, the probabilities are low, but differences in high *versus* low risk are apparent.

Care must be taken in interpreting maps such as these, because individual behavior has not been factored in. Background risk maps may therefore be most useful in developing the framework for a public education campaign to combat a disease endemic within a population or in developing a more in-depth epidemiological study. Knowledge of the explanatory variables used by the model to generate the risk map can be useful to identify the behaviors to target in such a campaign, for example warning residents of the risks associated with agricultural animal contact or playing in or around an impaired stream. In this way, limited public health resources can be targeted at the locations and the behaviors most associated with disease.

While geospatial analysis and the explicit inclusion of space when assessing disease risk can be a valuable tool to identify populations at risk, any model is only as good as the data and assumptions used to generate the model. Non-reporting of diseases because a patient does not seek medical attention, because a sample is not taken during the medical examination or due to incomplete reporting of disease incidence can erode the quality of a dataset. Patterns in non-reporting can also affect results if members of one socioeconomic group are less likely (or more likely) to be exposed or to seek treatment [46]. Care must be taken, therefore, in the interpretation of model results, because spatial differences in reporting rates may introduce bias into the modeled relationship between disease and environmental risk factors.

5. Conclusions

This research has shown that environmental variables are important risk factors for cryptosporidiosis and *E. coli* O157 infection in the individual. Proximity to karst geology was associated with both diseases, indicating that geology, specifically limestone and dolomite formations that are prone to karst weathering, should be incorporated as a proximity measure in waterborne disease risk models. Proximity to impaired surface water was a risk factor for cryptosporidiosis, but not *E. coli* O157 infection, likely related to the utility of *E. coli* as an indicator for *E. coli* O157.

Beef cow population density was positively associated with both diseases, while dairy cow population density was negatively associated with *E. coli* O157 infection. Because of the different results for cattle, populations should be partitioned into dairy and beef cattle when included as risk factors in a model for cryptosporidiosis and *E. coli* O157 infection.

Multiple transmission sources for both diseases are indicated by the positive association between percent agricultural land and percent developed land. Because both developed land and pasture land are associated with increased risk for cryptosporidiosis, future studies of environmental risk factors for

cryptosporidiosis should partition the cases by species (*C. hominis* and *C. parvum*) to examine the role of pasture and land use. To accomplish this, diagnostic methods must identify *Cryptosporidium* at the species level.

Models such as these can be useful to identify important risk factors for disease and to generate a background risk model. These results can then be used to develop a public education campaign to target limited public health resources to address behaviors associated with the most important risk factors in the communities where opportunities for those behaviors are most likely. These results can also be used as the framework upon which to develop a more comprehensive epidemiological or microbiological study to examine specific pathways for disease that focus on individual level risk factors.

Acknowledgments

The authors express thanks to Agricola Odoi (Department of Comparative Medicine, University of Tennessee) for assistance with epidemiology, to Nicholas Nagle and Yingkui Li (Department of Geography, University of Tennessee) for assistance with statistics and methodology and to Judy Manners (Tennessee Department of Health).

Author Contributions

Ingrid Luffman and Liem Tran conceived and designed the models; Ingrid Luffman performed the modeling; Ingrid Luffman and Liem Tran analyzed and interpreted the results; Ingrid Luffman wrote the paper.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Bunnell, J.E. Medical geology: Emerging discipline on the ecosystem-human health interface. *EcoHealth* **2004**, *1*, 15–18.
2. Dohoo, I.W.; Martin, W.; Stryhn, H. *Veterinary Epidemiologic Research*; AVC: Charlottetown, Canada, 2003.
3. Robertson, W.; Yasvinski, G. Exposure. In *Animal Waste, Water Quality and Human Health*; Dufour, A.P., Dufour, A., Eds.; IWA Publishing: London, UK, 2012; pp. 257–282.
4. Dura, G.; Pándics, T.; Kádár, M.; Krisztalovics, K.; Kiss, Z.; Bodnár, J.; Asztalos, A.; Papp, E. Environmental health aspects of drinking water-borne outbreak due to karst flooding: Case study. *J. Water Health* **2010**, *8*, 513–520.
5. Jiang, Y.; Wu, Y.; Groves, C.; Yuan, D.; Kambesis, P. Natural and anthropogenic factors affecting the groundwater quality in the Nandong karst underground river system in Yunan, China. *J. Contam. Hydrol.* **2009**, *109*, 49–61.
6. Boyer, D.G.; Pasquarell, G.C. *Agricultural Land Use Impacts on Bacterial Water Quality in a Karst Groundwater Aquifer*; Wiley Online Library: Hoboken, NJ, USA, 1999.

7. Kelly, W.R.; Panno, S.V.; Hackley, K.C.; Martinsek, A.T.; Krapac, I.G.; Weibel, C.P.; Storment, E.C. Bacteria contamination of groundwater in a mixed land-use karst region. *Water Qual. Expo. Health* **2009**, *1*, 69–78.
8. O'Reilly, C.E.; Bowen, A.B.; Perez, N.E.; Sarisky, J.P.; Shepherd, C.A.; Miller, M.D.; Hubbard, B.C.; Herring, M.; Buchanan, S.D.; Fitzgerald, C.C.; *et al.* A waterborne outbreak of gastroenteritis with multiple etiologies among resort island visitors and residents: Ohio, 2004. *Clin. Infect. Dis.* **2007**, *44*, 506–512.
9. Dong, B.-Q.; Yang, J.; Wang, X.-Y.; Gong, Y.; von Seidlein, L.; Wang, M.-L.; Lin, M.; Liao, H.-Z.; Ochiai, R.L.; Xu, Z.-Y.; *et al.* Trends and disease burden of enteric fever in Guangxi Province, China, 1994–2004. *Bull. World Health Organ*. **2010**, *88*, 689–696.
10. Rangel, J.M.; Sparling, P.H.; Crowe, C.; Griffin, P.M.; Swerdlow, D.L. Epidemiology of *Escherichia coli* O157:H7 outbreaks, United States, 1982–2002. *Emerg. Infect. Dis.* **2005**, *11*, 603–609.
11. Chalmers, R.M.; Aird, H.; Boldton, F.J. Waterborne *Escherichia coli* O157. *J. Aplied Microbiol. Symp. Suppl.* **2000**, *88*, 124S–132S.
12. Muniesa, M.; Jofre, J.; García-Aljaro, C.; Blanch, A.R. Occurrence of Escherichia coli O157:H7 and other enterohemorrhagic *Escherichia coli* in the environment. *Environ. Sci. Technol.* **2006**, *40*, 7141–7149.
13. O'Brien, S.J.; Goutam, K.A.; Gilham, C. Contact with farming environment as a major risk factor for Shiga Toxin (Vero Cytotoxin)-producing *Escherichia coli* O157 infection in humans. *Emerg. Infect. Dis.* **2001**, *7*, 1049–1051.
14. Strachan, N.J.; Dunn, G.M.; Locking, M.E.; Reid, T.M.; Ogden, I.D. *Escherichia coli* O157: Burger bug or environmental pathogen? *Int. J. Food Microbiol.* **2006**, *112*, 129–137.
15. Friesema, I.H.; van de Kassteele, J.; de Jager, C.M.; Heuvelink, A.E.; van Pelt, W. Geographical association between livestock density and human Shiga toxin-producing *Escherichia coli* O157 infections. *Epidemiol. Infect.* **2011**, *139*, 1081–1087.
16. Kassenborg, H.D.; Hedberg, C.W.; Hoekstra, M.; Evans, M.C.; Chin, A.E.; Marcus, R.; Vugia, D.J.; Smith, K.; Ahuja, S.D.; Slutsker, L.; *et al.* Farm visits and undercooked hamburgers as major risk factors for sporadic *Escherichia coli* O157:H7 infection: Data from a case-control study in 5 FoodNet sites. *Clin. Infect. Dis.* **2004**, *38*, S271–S278.
17. Klumb, C.; Robinson, T.; Cebelinski, E.; Alexander, B.H.; Smith, K. Evaluation of relationships between private well water, geologic sensitivity, cattle density, and *Cryptosporidium parvum* infection in Minnesota, 2000–2008. In Proceedings of the International Conference on Emerging Infectious Diseases, Minneapolis, MN, USA, 11–17 November 2010; pp. 669–670.
18. Hlavsa, M.C.; Brunkard, J.M. *Surveillance for Waterborne Disease Outbreaks and Other Health Events Associated with Recreational Water—United States, 2007–2008*; U.S. Department of Health and Human Services, Centers for Disease Control and Prevention: Atlanta, GA, USA, 2011.
19. Hlavsa, M.C.; Watson, J.C.; Beach, M.J. Cryptosporidiosis surveillance—United States 1999–2002. *MMWR Surveill. Summ.* **2005**, *54*, 1–8.
20. Michel, P.; Wilson, J.B.; Martin, S.W.; Clarke, R.C.; Mcewen, S.A.; Gyles, C.L. Temporal and geographical distributions of reported cases of *Escherichia coli* O157:H7 infection in Ontario. *Epidemiol. Infect.* **1999**, *122*, 193–200.

21. Valcour, J.E.; Michel, P.; McEwen, S.A.; Wilson, J.B. Associations between indicators of livestock farming intensity and incidence of human Shiga toxin-producing *Escherichia coli* infection. *Emerg. Infect. Dis.* **2002**, *8*, 252–257.
22. Scallan, E.; Hoekstra, R.M.; Angulo, F.J.; Tauxe, R.V.; Widdowson, M.-A.; Roy, S.-L.; Jones, J.L.; Griffin, P.M. Foodborne illness acquired in the United States—Major pathogens. *Emerg. Infect. Dis.* **2011**, *17*, 7–15.
23. Jarup, L. The Role of Geographical Studies in Risk Assessment. In *Spatial Epidemiology Methods and Applications*; Elliott, P., Wakefield, J., Best, N., Briggs, D., Eds.; Oxford University Press: New York, NY, USA, 2000; pp. 415–433.
24. Barton, K.E.; Howell, D.G.; Vigil, J.F. *The North America Tapestry of Time and Terrain*; Geologic Investigations Series I-2781; U.S. Geological Survey: Reston, VA, USA, 2003.
25. Godlberg, D.W.; Wilson, J.P. USC WebGIS Services. Available online: http://webgis.usc.edu/Services/Geocode/BatchProcess/Default.aspx (accessed on 22 November 2011).
26. Environmental Systems Research Institute, Inc. *ArcView 10.0 Desktop*; Environmental Systems Research Institute, Inc.: Redlands, CA, USA, 2011.
27. Greene, D.C.; Wolfe, W.J. 1:250,000 Geology of Tennessee. Available online: http://water.usgs.gov/GIS/metadata/usgswrd/XML/geo250k.xml#stdorder (accessed on 2 April 2011).
28. Denton, G.M.; Brame, C.J.; Arnwine, D.H.; Cartwright, L.K.; Graf, M.H. 2008 305(b) Report The Status of Water Quality in Tennessee. Available online: http://www.tn.gov/environment/water/docs/wpc/2008_305b.pdf (accessed on 2 April 2011).
29. Tennessee Department of Environment and Conservation. Final Year 2008 303(d) List. Available online: http://www.tn.gov/environment/water/docs/wpc/2008_303d.pdf (accessed on 2 April 2011).
30. EPA RADims. Available online: http://epamap32.epa.gov/radims/ (accessed on 17 October 2011).
31. United States Department of Agriculture. 2007 Census of Agriculture. Available online: http://quickstats.nass.usda.gov/?agg_level_desc=ZIP%20CODE (accessed on 2 April 2011).
32. Fry, J.; Xian, A.J.; Jin, S.; Dewitz, J.A.; Homer, C.G.; Yang, L.; Barnes, C.A.; Herold, N.D.; Wickham, J.D. Completion of the 2006 national land cover database for the conterminous United States. *Photogramm. Eng. Remote Sens*. **2011**, *77*, 858–864.
33. United States Census. 2010 Census Table B17017. Available online: http://www.census.gov/acs/www/data_documentation/summary_file/ (accessed on 19 April 2012).
34. Bivand, R.S.; Pebesma, E.J.; Gomez-Rubio, V. *Applied Spatial Data Analysis with R*; Gentleman, R., Hornik, K., Parmigiani, G., Eds.; Springer: New York, NY, USA, 2008.
35. R Core Team. *R: A Language and Environment for Statistical Computing Version 2.15.1*; R Foundation for Statistical Computing: Vienna, Austria, 2012.
36. Baddeley, A.; Turner, R. Spatstat: An R package for analyzing spatial point patterns. *J. Stat. Softw.* **2005**, *12*, 1–42.
37. Baddeley, A.; Berman, M.; Fisher, N.I.; Hardegen, A.; Milne, R.K.; Shah, S.R.; Turner, R. Spatial logistic regression and change-of-support in poisson point process. *Electron. J. Stat.* **2010**, *4*, 1151–1201.
38. Beach, M.J. *Waterborne: Recreational water*. In *Cryptosporidium and Cryptosporidiosis*, 2nd ed.; Fayer, R., Xiao, L., Eds.; CRC Press: Boca Raton, FL, USA, 2008; pp.329–362.

39. Hoek, M.R.; Oliver, I.; Barlow, M.; Heard, L.; Chalmers, R.; Paynter, S. Outbreak of *Cryptosporidium parvum* among children after a school excursion to an adventure farm, South West England. *J. Water Health* **2008**, *6*, 333–338.
40. Centers for Disease Control and Prevention. Cryptosporidiosis outbreak at a summer camp—North Carolina, 2009. *Morb. Mortal. Wkly. Rep.* **2011**, *60*, 918–922.
41. Robinson, J.A.; Brooks, J.M. *Public Water-Supply Systems and Associated Water Use in Tennessee*; Open-File Report 2010–1226; United States Geologic Survey: Reston, VA, USA, 2005.
42. Vesper, D.J.; White, W.B. Metal transport to karst springs during storm flow: An example from Fort Campbell, Kentucky/Tennessee, USA. *J. Hydrol.* **2003**, *276*, 20–36.
43. Wilkes, G.E.; Edge, T.; Gannon, V.; Jokinen, C.; Lyautey, E.; Medeiros, D.; Neumann, N.; Ruecker, N.; Topp, E.; Lapen, D.R. Seasonal relationships among indicator bacteria, pathogenic bacteria, *Cryptosporidium* oocysts, *Giardia* cysts, and hydrological indices for surface waters within an agricultural landscape. *Water Res.* **2009**, *43*, 2209–2223.
44. Dorner, S.M.; Anderson, W.B.; Gaulin, T.; Candon, H.L.; Slawson, R.M.; Payment, P.; Huck, P.M. Pathogen and indicator variability in a heavily impacted watershed. *J. Water Health* **2007**, *5*, 241–257.
45. Hunter, P.R.; Hadfield, S.J.; Wilkinson, D.; Lake, I.R.; Harrison, F.C.; Chalmers, R.M. Subtypes of *Cryptosporidium parvum* in humans and disease risk. *Emerg. Infect. Dis.* **2007**, *13*, 82–88.
46. Lake, I.R.; Harrison, F.C.; Chalmers, R.M.; Bentham, G.; Nichols, G.; Hunter, P.R.; Kovats, R.S.; Grundy, C. Case-control study of environmental and social factors influencing cryptosporidiosis. *Eur. J. Epidemiol.* **2007**, *22*, 805–811.
47. Kistemann, T.; Zimmer, S.; Vågsholm, I.; Andersson, Y. GIS-supported investigation of human EHEC and cattle VTEC O157 infections in Sweden: Geographical distribution, spatial variation and possible risk factors. *Epidemiol. Infect.* **2004**, *132*, 495–505.
48. Said, B.; Wright, F.; Nichols, G.L.; Reacher, M.; Rutter, M. Outbreaks of infectious disease associated with private drinking water supplies in England and Wales 1970–2000. *Epidemiol. Infect.* **2003**, *130*, 469–479.
49. Benham, B.L.; Affaut, C.; Zeckoski, R.W.; Mankin, K.R.; Pachepsky, Y.A.; Adeghi, A.M.; Rannan, K.M.; Oupir, M.L.; Habersack, M.J. Modeling bacteria fate and transport in watersheds to support TMDLs. *Am. Soc. Agric. Biol. Eng.* **2006**, *49*, 987–1002.
50. Dietz, V.J.; Roberts, J.M. National surveillance for infection with *Cryptosporidium parvum*, 1995–1998: What have we learned? *Public Health Rep.* **2000**, *115*, 358–363.
51. Suresh, K.G.; Toranzos, G.A.; Fayer, R.; Nissaparton, V.; Olveda, R.; Ashbolt, N.; Gannon, V. Assessing the importance of zoonotic waterborne pathogens. In *Animal Waste, Water Quality and Human Health*; Dufour, A.I., Bartram, J., Bos, R., Gannon, V., Eds.; IWA Publishing: London, UK, 2012; pp. 17–72.