Predicting Lyme Disease Incidence in Humans and Dogs

Katherine A. Cartagena Cody Hansen

> ISIB 2014 The University of Iowa College of Public Health

July 17th, 2014 Mentor: Dr. Kate Cowles

Katherine A. Cartagena Cody Hansen Predicting Lyme Disease Incidence in Humans and Do

Outline



Introduction

- Background Information
 - Definition of Lyme disease
 - Symptoms
 - Early Treatment
 - Tick Information



Data Sets

Variables

- Websites
- Variable Type
- Scatterplot Matrices



Models

Analysis and Conclusion

- Tables
- Maps
- Prediction Errors
- Study Limitations
- Conclusions and Future Work



References

Katherine A. Cartagena Cody Hansen

Predicting Lyme Disease Incidence in Humans and De

A B K A B K

Introduction

Research Goals:

- Predict new U.S. states where Lyme disease will occur.
- Determine whether inclusion of data on incidence in dogs improves the predictive ability of statistical models for human Lyme disease incidence.
- Identify the most significant predictor variables.

・ロト ・ 同ト ・ ヨト ・ ヨト ・

Definition of Lyme disease Symptoms Early Treatment Tick Information

Lyme Disease

- Infectious disease.
- Transmited by the bite of three different species belonging to the genus *Borrelia*.
- 36 48 hours or more before the disease can be transmitted.

Definition of Lyme disease Symptoms Early Treatment Tick Information

Symptoms of Lyme Disease

Early localized stage (3 - 30 days post tick bite)

- Erythema migrans
- Fatigue
- Fever
- Headache
- Muscle and Joint aches
- Swollen lymph nodes



Katherine A. Cartagena Cody Hansen

Predicting Lyme Disease Incidence in Humans and De

Definition of Lyme disease Symptoms Early Treatment Tick Information

Symptoms of Lyme Disease

Early disseminated stage (days to weeks post tick-bite)

- More EM lesions in other areas of the body
- Bell's palsy
- Swelling of large joints
- Due to meningitis, suffer from severe headaches and neck stiffness
- Changes in heartbeat

(日) (四) (日) (日)

Definition of Lyme disease Symptoms Early Treatment Tick Information

Why is it important to treat Lyme disease early?

Prevention of the following:

- Shooting pains
- Numbress or tingling in the hands or feet
- Problems with short-term memory



Definition of Lyme disease Symptoms Early Treatment **Tick Information**

The Ticks

- Blacklegged Tick (*Ixodes* scapularis)
- Western Blacklegged Tick (*Ixodes pacificus*)





< ロト (四) (三) (三)

Definition of Lyme disease Symptoms Early Treatment **Tick Information**



・ロト ・ 理ト ・ ヨト ・ ヨト

3

Definition of Lyme disease Symptoms Early Treatment **Tick Information**

Life Cycle of Ticks

- Their life cycle typically lasts two years, comprising of 4 life stages: egg, larva, nymph and adult.
- At every life stage after egg, the tick must feed.



Predicting Lyme Disease Incidence in Humans and Do

Katherine A. Cartagena Cody Hansen

Definition of Lyme disease Symptoms Early Treatment **Tick Information**

Infection Source

- Ticks are not born infected with *Borrelia burgdorferi*.
- White-footed mice (*Peromyscus leucopus*) is the primary reservoir.



< 日 > < 同 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ >

Definition of Lyme disease Symptoms Early Treatment **Tick Information**

Nymphs

- Nymphs seek larger hosts
- Commonly found in near ground vegetation



Katherine A. Cartagena Cody Hansen

Predicting Lyme Disease Incidence in Humans and Do

Definition of Lyme disease Symptoms Early Treatment **Tick Information**

- Humans acquire the infection through Nymphs
- Takes 36 to 48 Hours
- Adult ticks can also infect hosts



Definition of Lyme disease Symptoms Early Treatment **Tick Information**

- Hard to detect infection
- Often found during yearly testing
- Canary?



- - E - F -

Data Sets for Model Fitting:

- Human Lyme Incidence from 2007 2011
- Dog Incidence Rates from 2011

Data Sets for Prediction:

- Human Lyme Incidence from 2012
- Dog Incidence Rates from 2012

▲ □ ▶ ▲ □ ▶ ▲ □ ▶

Websites Variable Type Scatterplot Matrices

General Predictor Variables

1 Temperature

www.esrl.noaa.gov/psd/data/usclimate/tmp.state.19712000.climo
Humidity

www.currentresults.com/Weather/US/humidity-by-state-in-summer.Php

3 Land Usage

www.ers.usda.gov/data-products/major-land-uses.aspx#.U680qx_HnVM

4 Elevation

en.wikipedia.org/wiki/List_of_U.S._states_by_elevation

- **5** Dog Incidence Rates
 - www.capcvet.org/parasite-prevalence-maps/

Websites Variable Type Scatterplot Matrices

Variables

- Response Variable:
 - State Total
- Predictor Variables:
 - Dogs Incidence
 - ► Forest
 - Urban
 - Morning and Afternoon Humidity
 - January and July Temperatures
 - Deer Harvest
 - Mean and Highest Elevation

・ 同 ト ・ ヨ ト ・ ヨ ト

Websites Variable Type Scatterplot Matrices

Model Variables



Katherine A. Cartagena Cody Hansen

Predicting Lyme Disease Incidence in Humans and Do

Websites Variable Type Scatterplot Matrices

Model Variables



Katherine A. Cartagena Cody Hansen

Predicting Lyme Disease Incidence in Humans and Do

glm and hglm Models

- glm
 - Poisson regression model.
 - Ignores spatial autocorrelation in response variable, so significance of predictor variables is likely to be overestimated.
- hglm
 - Hierarchical poisson regression with a random effect for each state.
 - Conditional autoregressive model for random effects to capture spatial autocorrelation.
 - Empirical Bayes rather than fully Bayes model; MCMC was not used to fit.

イロト イポト イヨト イヨト

3

glm and hglm Models

▶ glm model:

$$\begin{split} log(E(Y_i)) &= \beta_0 + \beta_1(DogsInci_i) + \beta_2(ForestProp_i) + \\ \beta_3(UrbanProp_i) + \beta_4(MorningHum_i) + \beta_5(AfternoonHum_i) + \\ \beta_6(JanTemp_i) + \beta_7(JulyTemp_i) + \beta_8(DeerHarvest_i) + \\ \beta_9(MeanElev_i) + \beta_{10}(HighestElev_i) + log(E_i) \end{split}$$

▶ hglm model:

 $log(E(Y_i)) = \beta_0 + \beta_1 (JanTemp_i) + \beta_2 (JulyTemp_i) + \beta_3 (DogsInci_i) + log(E_i) + \phi_i$

▲□▶ ▲□▶ ▲□▶ ▲□▶ □ のの⊙

Tables Maps Prediction Errors Study Limitations Conclusions and Future Work

glm Tables of Coefficients and P-Values

GLM	Estimate Std.	Std. Error	z-value	Pr(> z)
(Intercept)	-5.34295	0.31337	-17.049	$< 2e^{-}16$ ***
DogsInci	35.15061	0.23303	150.838	$< 2e^{-16}$ ***
ForestProp	-3.04855	0.05568	-54.744	$< 2e^{-16}$ ***
UrbanProp	-5.22102	0.08494	-61.462	$< 2e^{-16}$ ***
MorningHum	0.08860	0.00125	70.369	$< 2e^{-16}$ ***
AfternoonHum	-0.01621	0.00125	-12.934	$< 2e^{-16}$ ***
JanTemp	-0.06322	0.00186	-33.879	$< 2e^{-16}$ ***
JulyTemp	0.01028	0.00422	2.432	0.015 *
DeerHarvest	0.01362	0.00045	29.794	$< 2e^{-16}$ ***
MeanElev	-1.17139	0.02733	-42.854	$< 2e^{-16}$ ***
HighestElev	-1.79525	0.03726	-48.177	$< 2e^{-16}$ ***

Katherine A. Cartagena Cody Hansen Predicting Lyme Disease Incidence in Humans and De

→ 3 → 4 3

Tables Maps Prediction Errors Study Limitations Conclusions and Future Work

hglm Tables of Coefficients and P-Values

HGLM	Estimate Std.	Std. Error	t-value	Pr(> t)
(Intercept)	-10.55877	3.32192	-3.179	0.0864 .
DogsInci	29.22464	4.73341	6.174	0.0252 *
JanTemp	-0.10898	0.02742	-3.975	0.0579 .
JulyTemp	0.15415	0.05440	2.833	0.1053

Tables Maps Prediction Errors Study Limitations Conclusions and Future Work

Average 2007 - 2011 State Counts



Tables Maps Prediction Errors Study Limitations Conclusions and Future Work

Predicted glm 2012 State Counts vs. Observed



Predicted GLM 2012 State Counts

Observed 2012 State Counts



< ロト < 同ト < ヨト < ヨト

Tables Maps Prediction Errors Study Limitations Conclusions and Future Work

Predicted hglm 2012 State Counts vs. Observed



Predicted HGLM 2012 State Counts

Observed 2012 State Counts



< ロト < 同ト < ヨト < ヨト

Tables Maps **Prediction Errors** Study Limitations Conclusions and Future Work

Mean Squared Prediction Errors for 2012 Counts

Out of sample prediction using GLM model:

Model including dog incidence (Raw Counts): mean((Actual2012Counts - Predicted2012CountsDogs) ²)	249598
Model omitting dog incidence (Raw Counts): mean((Actual2012Counts - Predicted2012CountsNoDogs) ²)	402794.3
Model including dog incidence (log Transformation): mean((log(Actual2012Counts) - log(Predicted2012CountsDogs)) ²)	11.00701
Model omitting dog incidence (log Transformation): mean((log(Actual2012Counts) - log(Predicted2012CountsNoDogs)) ²)	10.4362

(日) (四) (日) (日)

Tables Maps Prediction Errors **Study Limitations** Conclusions and Future Work

Study Limitations

- Lyme disease cases are tabulated by state of residence, not state of exposure.
- Not every case of Lyme disease is reported to CDC.
- All the variables are aggregated by state.
- No new states with Lyme disease incidence were found.
- 2012 cases were recorded differently.

< ロト (四) (三) (三)

Tables Maps Prediction Errors Study Limitations Conclusions and Future Work

Conclusions and Future Work

- hglm's predictions are much closer than glm's.
- Dog incidence rate is the most highly significant predictor in sample.
- January and July Temperatures were highly significant predictors.

Future Work:

- ▶ Report Models and Answers to the CDC.
- Write a paper for submission in an applied statistics journal.
- ▶ Predicting 2013 when incidence rates are available.

Katherine A. Cartagena Cody Hansen Predicting Lyme Disease Incidence in Humans and Do

References

- 1 http://www.tickencounter.org/images/Ixodes_scapularis.jpg
- 2 www.cdc.gov/ticks/images/lgmap-blacklegged_tick.jpg
- 3 www.cdc.gov/ticks/images/lgmap-western_blacklegged_tick.jpg
- 4 www.cdc.gov/ticks/images/lifecycle.jpg
- 5 www.dcnature.com/photosfull/Whitefooted%20Mouse%20(Peromyscus%20 leucopus).jpg
- 6 www.factmonster.com/ipka/A0004986.html

▲ □ ▶ ▲ □ ▶ ▲ □ ▶

Citation for R

@Manual, title = R: A Language and Environment for Statistical Computing, author = R Core Team, organization = R Foundation for Statistical Computing, address = Vienna, Austria, year = 2014, url = http://www.R-project.org/,

• • = • • = •

Citation for R packages

- @Article, title = hglm: A Package for Fitting Hierarchical Generalized Linear Models, author = Lars Ronnegard and Xia Shen and Moudud Alam, journal = The R Journal, year = 2010, volume = 2, number = 2, pages = 20-28, url = http://journal.r-project.org/archive/2010-2/RJournal_2010-2_Roennegaard et al.pdf,
- @Manual, title = maps: Draw Geographical Maps, author = Original S code by Richard A. Becker and Allan R. Wilks. R version by Ray Brownrigg. Enhancements by Thomas P Minka jtpminka@media.mit.edu¿, year = 2014, note = R package version 2.3-7, url = http://CRAN.R-project.org/package =maps,